# The Processing of Non-Nominal Metaphors

Dissertation

zur Erlangung des akademischen Grades

*Doktor der Philosophie (Dr. phil.)*

Eingereicht an der Sprach- und literaturwissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

von

**Camilo Rodríguez Ronderos, M.Sc.**

Gutachter/Innen:

Prof. Dr. Pia Knoeferle

Prof. Dr. Ernesto Guerra

Prof. Dr. Rasha Abdel Rahman

Datum der Disputation: 26. Januar 2021

Prof. Dr. -Ing. Dr. Sabine Kunst
Präsidentin der Humboldt-Universität zu Berlin

Prof. Dr. Stefan Kipf
Dekan der Sprach- und literaturwissenschaftlichen Fakultät

# Acknowledgements

I have always enjoyed the genre of 'acknowledgment sections of doctoral theses'. This is partly because, when I was still an undergrad, they struck me as being somewhat overly dramatic. Why do these researchers have such intense feelings of gratitude, and why do they have the urge to share words of wisdom about the thesis-writing process with their readers? I have a bit more clarity on this issue now that I am reaching the end of my own dissertation: It takes a village to raise a linguist, and without the enormous professional and emotional support that I have received from so many people, I would have not even come close to finishing this work.

The first person I want to thank is Pia Knoeferle, my advisor, who guided me through this process. She has always been deeply invested both in my education as a psycholinguist and in improving the overall quality of my work and my writing, and I can't thank her enough for everything she has taught me. I consider myself extremely lucky to have had the opportunity to work with her over the past four years. I also want to thank my second advisor, Ernesto Guerra, for his important assistance on this and in other projects. Ernesto is one the kindest people I have ever met. He has patiently helped me out whenever it was necessary, and I cherish the fact that we could spend hours and hours discussing anything about work or about life in general. I also want to thank Rasha Abdel Rahman, who served as external reviewer of this dissertation, for her encouragement and her helpful comments on this work.

My current and past lab mates have been a special group of people and I am grateful for all their help. They have, on more than one occasion, convinced me not to quit my PhD and get a job at the burrito stand on Friedrichstraße instead: Katja Maquate, Daniela Palleschi, Thi Thu Huong Nguyen, Aine Ito, Dato Abashidze and Alba Rodríguez. Time will tell if they made the right decision. I am also grateful to the student assistants who helped me create materials for some of the experiments in this dissertation and assisted me with recruiting and testing participants in the lab: Esma Tanis, Lea-Sophie Adam, Jakob Wünsch, Emanuelle Borchert, Thao Tran, Melis Odabaş and Olga Buchmüller.

I am very thankful to a lot of people in the Xprag, psycholinguistic and broader scientific communities who have given me valuable feedback on several stages of this project, both from a theoretical and experimental perspective: Valentina Bambini, Filippo Domaneschi, Ray Gibbs, Deirdre Wilson, Dedre Gentner, Sol Lago, Umesh

Patil, Dale Barr and Judith Degen. Special thanks to Ingrid Lossius Falkum, Petra Schumacher and Joseph DeVeaugh-Geiss, for inviting me to their universities to speak about different stages of this dissertation. This resulted in productive conversations that helped me give shape to this project.

Before my time at the HU, I worked as a student assistant at the Zentrum für Allgemeine Sprachwissenschaft (ZAS). There, I was exposed to people and ideas that had a big impact on me. In particular, I want to thank Jack Tomlinson, who gave me my first job in academia, Stefanie Solt, who taught the first class I ever took on semantics and pragmatics, Manfred Krifka, Uli Sauerland and Fabienne Salfner, who provided support and encouragement for me to take part in various conferences and events during my time at ZAS, and Bob van Tiel, Guillermo del Pinal and Judith Tonhauser, who often engaged with my ideas and pointed me in the direction of relevant literature. It was because of working at ZAS that I became passionate about using experiments to test theories of figurative language comprehension.

I want to thank Ira Noveck for being an amazing mentor and for providing great support throughout the years. Ira has deeply influenced my thinking and the way that I approach my work, and he has always had words of encouragement and advice when I needed them most. Nicole Gotzner is another extraordinary person to whom I owe quite a lot, and she has helped me navigate the hardships of being a PhD student more than anyone. Nicole and Ira are my role models for what I aspire to be both as a scientist and as person.

This dissertation would have been an insurmountable task without the following software: Ibex Farm (Drummond, 2013) and the Penncontroller extension (Zehr & Schwarz, 2018); R (R Core Team, 2020) RStudio (RStudio Team, 2020) and the R-packages MASS (Venables & Ripley, 2002), citr (Aust, 2019), dplyr (Wickham et al., 2020), ggplot2 (Wickham, 2016), here (Müller, 2017), knitr (Xie, 2015), lme4 (Bates et al., 2015), rmarkdown (Allaire et al., 2020), SimR(Green & MacLeod, 2016), stargazer (Hlavac, 2018), tidyr (Wickham & Henry, 2020) and xtable (Dahl et al., 2019). Particularly helpful were the R-Package VWPre (Porretta et al., 2017), which is so good that it almost made me enjoy pre-processing Visual World Data, and the rmarkdown template oxforddown (Lyngs, 2019), which made it possible for me to write this entire dissertation in R Markdown.

One thing that I discovered during this time is that there is a useful side-effect to having really smart, kind and generous friends: You can exploit their talents for the benefit of your work! Charlie, Vera, Mareike, Daniela, Alex, Biene, Stella, Heather, Sol, Umesh, Antonio, Matt, Mariana and many others have helped me pilot my experiments and find participants for them, they have given their linguistic intuitions or their statistical advice on various topics related to this dissertation and they have heard me go on and on about metaphors for the longest time without complaining. They've also been, together with other friends and family, a vital support network, for which I am very thankful. Without them I wouldn't have survived this phase of my life, especially given that the last stages of my dissertation coincided with the outbreak of a world pandemic.

Finally and most importantly, I want to thank Elinor for all the support she has given me throughout the last six years. There are no metaphors good enough to describe how incredible she is. This thesis is dedicated to her.

Camilo R. R.
Kreuzberg, Berlin
11 August 2020

# Abstract

Metaphors are a puzzling phenomenon of language use. Why is it that we can understand when a word such as *princess* is used in a way that deviates dramatically from its de-contextualized conventional meaning in a sentence such as *my cat is a princess*?

Two main sets of theories attempt to answer this question by positing the involvement of different cognitive mechanisms during processing. The first one, which I refer to as the Implicit Comparison View, claims that metaphors are understood through a process of analogical reasoning in which the elements of a metaphoric expression (in the example above *my cat*, which is known as the 'topic' and *princess*, which is known as the 'vehicle') are scanned for relational similarities, i.e. ways in which the internal structure of the metaphoric elements is coherent across elements (e.g. Gentner et al., 2001; Gentner & Bowdle, 2008). Once the structure of both elements has been aligned, inferences are projected from vehicle to topic in order to reach a final utterance interpretation.

A second view, which I refer to as the Category Inclusion View, sees metaphor comprehension as a process in which the lexical meaning of the metaphoric vehicle is spontaneously changed to represent a newly created, goal-oriented category (e.g. Glucksberg, 2008; Sperber & Wilson, 2008). To understand a metaphor, a listener must adjust the meaning of the vehicle given the parameters set by the metaphoric topic together with the relevant context.

Despite there being a large body of experimental data testing the predictions made by these theories (e.g. Bowdle & Gentner, 2005; Gernsbacher et al., 2001; Jones & Estes, 2005; Jones & Estes, 2006; McGlone & Manfredi, 2001; Wolff & Gentner, 2011), it has not been possible to settle this debate and tip the scale in favor of one or the other view (see Holyoak & Stamenković, 2018, for a systematic review of the empirical studies). Holyoak and Stamenković (2018) mention two ways in which the existing debate on metaphor processing could be moved forward, namely by examining different types of metaphors other than so-called nominal metaphors (i.e. metaphors of the type 'X is a Y') and by examining metaphors in languages other than English.

This dissertation attempts to do just that by examining the processing of two types of German non-nominal metaphors: Verbal metaphors and verb-object

metaphors. This was done by investigating the role of context during metaphor comprehension in order to further specify the available theories, and, more generally, by drawing on the literature on situated and incremental language processing (see Huettig et al., 2011; Huettig et al., 2012; Kamide, 2008; Knoeferle & Guerra, 2016, for reviews).

Specifically, I address three issues for which the available theories make opposing predictions: (1) The role of the literal meaning of the metaphoric vehicle during processing, (2) the (a)symmetry of the metaphoric elements and (3) the effect of metaphor conventionality, familiarity and aptness during processing. I investigated these issues across 14 Experiments making use of Eye-Tracking during reading, Eye-tracking during concurrent processing of spoken language and visual context (known as the Visual World Paradigm), reaction-time and self-paced reading measures as well as offline rating tasks.

The results in relation to issue (1) suggest that, during processing of verbal metaphors (such as *The journalist's opinion was fenced-in after the change in regime*), features associated exclusively with the literal meaning of the metaphoric verb (the feature of physical containment) neither facilitate nor hinder processing, despite facilitating processing of the same verbs when these were presented without a sentential context. Regarding issue (2), the studies show that verb-object metaphors (such as *Sebastian feeds a princess*, uttered when Sebastian is feeding his cat) are not processed symmetrically: When the verbal topic *feeds* appears before the vehicle *princess*, participants settled on a metaphoric interpretation more swiftly upon hearing the vehicle *princess* than when they heard the vehicle before the topic. Finally, regarding issue (3), metaphoric aptness (defined as the degree to which the figurative meaning of a metaphoric vehicle captures relevant features of the metaphoric topic) was shown to facilitate processing of both verbal and verb-object metaphors, whereas conventionality (defined as the frequency with which a metaphoric vehicle is used in its metaphoric meaning) and familiarity (defined as the frequency of a specific topic-vehicle pair) had no such effect.

Overall, I interpret the results as being more consistent with the Category Inclusion View than with the Indirect Comparison View in each one of the three issues investigated. The current dissertation thus makes an important contribution towards resolving the debate on the cognitive mechanisms involved during metaphor comprehension.

# Deutsche Zusammenfassung

Metaphern sind ein rätselhaftes Phänomen des Sprachgebrauchs. Warum sind wir in der Lage zu verstehen, wenn ein Wort wie *Prinzessin* auf einer Weise verwendet wird, die dramatisch von der de-kontextualisierten, konventionellen Bedeutung des Wortes abweicht, wie z.B. in dem Satz *Meine Katze ist eine Prinzessin*? Zwei Theorien versuchen, diese Frage zu beantworten, indem sie die Beteiligung verschiedener kognitiver Mechanismen während der Verarbeitung postulieren. Die erste, die ich als "Implicit Comparison View" bezeichne, behauptet, dass Metaphern durch einen Prozess des analogen Denkens verstanden werden. Als Teil dieses Prozesses werden die Elemente eines metaphorischen Ausdrucks (im obigen Beispiel *meine Katze*, und *Prinzessin*, die jeweils "metaphorisches Tenor" und "metaphorisches Vehikel" genannt werden) auf relationale Ähnlichkeiten geprüft, d. h., auf die Arten, in denen die internen Strukturen der metaphorischen Elemente miteinander kohärent sind (z.B. Gentner et al., 2001; Gentner & Bowdle, 2008). Sobald die maximale Kohärenz zwischen den Strukturen beider Elemente etabliert wurde, werden Inferenzen vom Vehikel zum Tenor projiziert, um eine endgültige Interpretation der Äußerung zu erreichen.

Eine zweite Ansicht, die ich als "Category Inclusion View" bezeichne, sieht das Verstehen einer Metapher als einen Prozess, bei dem die lexikalische Bedeutung des metaphorischen Vehikels spontan moduliert wird, um eine ad-hoc, zielorientierte Kategorie zu schaffen (z.B. Glucksberg, 2008; Sperber & Wilson, 2008). Um eine Metapher zu verstehen, muss die Bedeutung des Vehikels unter Berücksichtigung der vom metaphorischen Tenor festgelegten Parameter kontextuell angepasst werden. Obwohl es eine große Anzahl an Experiment gibt, die die Vorhersagen dieser beiden Theorien testen (z.B. Bowdle & Gentner, 2005; Gernsbacher et al., 2001; Jones & Estes, 2005; Jones & Estes, 2006; McGlone & Manfredi, 2001; Wolff & Gentner, 2011) ist es bis jetzt nicht möglich gewesen, das Problem der Metaphernverarbeitung zu lösen zugunsten einer der beiden Theorien (siehe Holyoak & Stamenković, 2018, für eine systematische Überprüfung der empirischen Studien). Holyoak and Stamenković (2018) erwähnen zwei Möglichkeiten, wie die bestehende Debatte über die Verarbeitung von Metaphern vorangebracht werden könnte, nämlich (i) durch das Untersuchen verschiedener Arten von Metaphern jenseits der sogenannten nominalen Metaphern (d. h. Metaphern vom Typ *X ist ein Y*), und (ii) durch das

Untersuchen von Metaphern in anderen Sprachen als Englisch. Diese Dissertation versucht genau das zu tun, indem die Verarbeitung von zwei Arten deutscher nicht-nominaler Metaphern untersucht werden: verbale Metaphern und Verb-Objekt-Metaphern. Dies wurde gemacht durch eine Untersuchung der Rolle des Kontexts während der Verarbeitung von nicht-nominalen Metaphern. Dabei wurde auf die Literatur zur situierten und inkrementellen Sprachverarbeitung zurückgegriffen (siehe Huettig et al., 2011; Huettig et al., 2012; Kamide, 2008; Knoeferle & Guerra, 2016). Insbesondere spreche ich drei Themen an, für die die verfügbaren Theorien entgegengesetzte Vorhersagen treffen: (1) Die Rolle der wörtlichen Bedeutung des metaphorischen Vehikels während der Verarbeitung, (2) die (a)Symmetrie der Reihenfolge der metaphorischen Elemente und (3) die Effekte von Konventionalität, Bekanntheitsgrad und Angemessenheit der einzelnen Metaphern auf die Verarbeitung. Ich habe diese Themen in 14 verschiedenen Experimenten untersucht, bei denen Reaktions- und Lesezeiten, Augenbewegungen sowie Offline-Bewertungen gemessen wurden. Die Ergebnisse in Bezug auf (1) legen nahe, dass während der Verarbeitung verbaler Metaphern (wie z.B. *die Meinung des Journalisten wurde nach dem Regimewechsel eingezäunt*), Merkmale, die ausschließlich mit der wörtlichen Bedeutung des metaphorischen Verbs verbunden sind, die Verarbeitung der Metaphern weder erleichtern noch verhindern, obwohl diese Merkmale die Verarbeitung derselben Verben erleichtern, wenn die Verben ohne Satzkontext präsentiert werden. In Bezug auf (2) zeigen die Studien, dass Verb-Objekt-Metaphern (wie z.B. *Sebastian füttert eine Prinzessin*, wenn Sebastian seine Katze füttert) asymmetrisch verarbeitet werden: Wenn das verbale Tenor (z.B. *füttert*) vor dem Vehikel (z.B. *Prinzessin*) gehört wurde, war es einfacher für die Teilnehmer des Experiments sich für eine metaphorische Interpretation der Äußerung beim Hören des Vehikels zu entscheiden als wenn sie das Vehikel vor dem Tenor gehört haben. In Bezug auf (3), zeigen die Ergebnisse der Experimente, dass die Angemessenheit einer Metapher (d.h., inwiefern die übertragene Bedeutung eines Vehikels relevante Merkmale vom Tenor erfasst) einen größeren Einfluss auf die Verarbeitung von verbalen und verb-objekt Metaphern hatte als Konventionalität (d.h. die Häufigkeit, mit der ein metaphorisches Vehikel in seiner metaphorischen Bedeutung verwendet wird) und Bekanntheitsgrad (d.h. die Häufigkeit eines bestimmten metaphorischen Tenor-Vehikel-Paares). Insgesamt interpretiere ich die Ergebnisse als besser zu vereinbaren mit der “Category Inclusion View” als mit der “Implicit Comparison View” in jedem der drei untersuchten Themen. Diese Dissertation leistet somit einen wichtigen Beitrag, um die Debatte über die kognitiven Mechanismen, die beim Verständnis von Metaphern beteiligt sind, voranzubringen.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

**ASL** . . . . . . American Sign Language

**CMT** . . . . . Conceptual Metaphor Theory

**ERP** . . . . . . Event-related Potentials

**SPM** . . . . . . Standard Pragmatic Model

**SME** . . . . . . Structure Mapping Engine

**VWP** . . . . . Visual World Paradigm

# 1 Introduction

A fundamental property of language is that the meaning of words is predominantly stable. This is an obvious fact to any language user: When navigating through a city, for example, it is necessary to know the meaning of the words we read on street signs. We rely on the fact that words such as “stop”, “school”, “danger”, etc. have a public, conventional meaning that is in line with the meaning those words have in our heads, and this allows us to act accordingly. If we do not know the meaning of one of these words, we can consult a dictionary, which lists the conventional meaning(s) of the words in a language. If the meaning of words were not stable, there would be no dictionaries. This layman’s notion of the stability of the meaning of words is shared by different theories of linguistic meaning. Philosophers of language, for example, have claimed that understanding the meaning of a sentence is equivalent to understanding the meaning of the individual words and the combinatorial processes through which these words are associated (Davidson, 2001). In this sense, the meaning of any given word is assumed to be stable. This is taken as an axiom of the construction of linguistic meaning, which is best exemplified in Jerry Fodor’s “disquotational” view of the lexicon (Fodor, 1998, p. 56; Fodor & Lepore, 2002, p. 95). According to Fodor, word meaning is non-decomposable. It can only be identified by its denotation: The meaning of the word *keep* is “keep”, the meaning of the word *house* is “house”, and so forth.

(Psycho)Linguists tend to share a similar view, generally stating that a lexical unit is a stable bundle of phonological, semantic and syntactic information stored together as an entry in a mental dictionary that contains the senses of all the words we know, known as the mental lexicon (Jackendoff, 2002, p. 130-131; Johnson-Laird, 1987; Marslen-Wilson, 1987; i.a., but see Elman, 2009, for a theory of lexical meaning without a mental lexicon). This mental lexicon is assumed to be shared by the members of a language community via tacit agreement (Clark, 1996), which allows us to trust that our interlocutors will know the meaning of these words when we use them. To go beyond the meaning of words and understand what a sentence means, a language user has to access the meaning of the necessary words stored in the mental lexicon and then put them together via semantic and syntactic rules (see Chierchia & McConnell-Ginet, 2000).

There are good reasons to believe that this view is an accurate description of how we understand language. Empirical evidence has shown that members of a language community rapidly converge on using identical words to refer to the same novel entities (Clark & Wilkes-Gibbs, 1986; Garrod & Doherty, 1994; Markman & Makin, 1998), lending support to the idea that language users do in fact use the same words to express very similar concepts. Furthermore, studies have reliably shown that language users exploit their knowledge about the properties of individual entries in their mental lexicon to anticipate upcoming information, such as thematic role knowledge (Altmann & Kamide, 1999; Taraban & McClelland, 1988; Trueswell et al., 1994). These anticipatory effects are only plausible under the assumption that language users reliably access a conceptual representation of the words they hear or read, and that this representation is stable across participants in an experiment.

However, when assuming the perspective of stable lexical units that are readily retrieved to construct meaning, some linguistic phenomena become very hard to account for. Such is the case of metaphors. Metaphors are instances of language use in which words from very different conceptual domains are combined in a way in which they seem to spontaneously create new meaning that is different from what the individual words typically convey. Take sentences 1-3 below:

*1) Some lawyers are sharks.*

*2a) Sebastian is feeding a princess.*

*2b) Sebastian is feeding a princess. [uttered only in the context of Sebastian feeding his spoiled cat]*

*3) The journalist's opinion was fenced in after the change in regime.*

In these examples, some words are used in a way such that the intended meaning of a speaker is incompatible with the word's base-level meaning, yet a meaningful overall interpretation of the expression is not only possible, but readily available. In (1), the speaker putatively does not mean to say that some lawyers are large marine predators, but instead intends to say that a subset of lawyers is particularly vicious and aggressive. This type of metaphoric expression is known as a nominal metaphor, and makes up the core of psycholinguistic investigations on metaphor understanding.

Examples (2a) and (2b) show a crucial aspect of metaphor processing, namely its context sensitivity. Unlike in (1), the speaker of (2a) might be using the word *princess* in its conventional sense (If, for example, Sebastian is feeding a member of a royal family). However, when the same sentence is uttered in a constraining context, as is the case in (2b), it becomes clear that *princess* is being used metaphorically to communicate something about Sebastian's cat that goes beyond (and might be totally incompatible with) the conventional meaning of *princess* (perhaps simply that the cat is spoiled). The difference between (2a) and (2b) shows that context alone can trigger dramatically different interpretations based on the way in which individual words are understood. This highlights the fact that metaphoric meaning doesn't just come about as the result of the interaction between two words (*lawyers* and *sharks*, in example (1)), but is also the product of how the meaning of words is affected by the context in which they are embedded.

Finally, it is not just nouns that can be used metaphorically. In (3), the verb *fenced in*, which typically entails physical containment, is used to predicate over the noun *opinion*, which does not have a physical dimension at all. The speaker of (3) probably means to communicate something different, namely that the journalist was not able to freely speak her mind (or a related proposition).

Examples 1-3 can be described as novel metaphors: Their metaphoric meaning is not conventionally associated with any specific word: it must be computed on the spot as opposed to being retrieved, which would instead be the case of fossilized metaphors such as the word *leg* in the expression *The leg of the table*.

Keeping the mental lexicon view in mind, how is it that we are able to derive meaning from sentences 1-3 with relative ease despite the fact that the words are being used in such an unconventional way? This question is essential in metaphor processing research.

Specifically, metaphor research is concerned with trying to make sense of the following three apparent facts about language:

(a) words have relatively stable conventional meanings

(b) metaphors are pervasive in everyday communication and in languages across the world (Glucksberg, 1989; Lakoff & Johnson, 2008)

(c) metaphoric language is similarly effortful to understand compared to non-metaphoric language (Gibbs, 2002)

These three facts, taken together, represent somewhat of a puzzle of metaphoric meaning. We rely on words having stable meanings so that we can efficiently communicate with one another. However, metaphoric meaning comes about by spontaneously and dramatically changing this conventional meaning, and yet, it seems to be a widespread phenomenon that need not come at a cost for the listener to comprehend. How is this possible? Answering this question is paramount if we want to understand how linguistic meaning is created in our minds, associated with individual words and reliably shared within language communities.

The past 30 years of psycholinguistic research have brought about extensive work trying to answer this question (see the articles in Gibbs, 2008). Still, we are far from reaching a consensus on even the most basic cognitive mechanisms involved in the comprehension of a novel metaphor (Gentner & Bowdle, 2008; Glucksberg, 2008; Tendahl & Gibbs, 2008; Wilson, 2011).

In terms of stipulating concrete cognitive mechanisms at play, the theoretical views on metaphor processing can be broadly split into two sets of views. The first one, which I'll refer to as the Implicit Comparison view, claims that metaphor comprehension takes place through a process of analogical reasoning in which the elements of a metaphoric expression (in sentence (1) *some lawyers*, which is known as the **metaphoric topic** and *sharks*, which is known as the **metaphoric vehicle**) are scanned for relational similarities, i.e. ways in which the internal structure of the metaphoric elements is coherent across elements (Boroditsky, 2000; Coulson & Oakley, 2005; Gentner & Bowdle, 2008; Gentner et al., 2001; Thibodeau & Durgin, 2011; Wolff & Gentner, 2011). In this view, the puzzle of word meaning is solved by assuming that the stable conventional meaning of all individual words is retrieved as it normally would and the metaphoric meaning of the expression is understood by establishing systematic relationships between these meanings, as one would do in an explicit comparison ("Some lawyers are like sharks in some relevant way").

A second view, the Category Inclusion view, sees metaphor comprehension as a process in which the lexical meaning of the metaphoric vehicle is spontaneously changed to represent a newly created, goal-oriented category (McGlone & Manfredi, 2001; Rubio Fernandez, 2007; Wilson & Sperber, 2012). A listener does not look for ways in which the internal structure of topic and vehicle are similar to one another. Instead, the meaning of the vehicle changes by either acquiring dual-reference (Glucksberg, 2008) or through broadening and narrowing of the meaning stored in the mental lexicon (Sperber & Wilson, 2008). The metaphoric topic is then understood to be a member of this category and the sentence as a whole is understood much like a regular Category Inclusion statement such as *Apples are fruits*. This way, the meaning puzzle is solved by positing that the conventional meaning of the metaphoric vehicle is rapidly modulated given a set of parameters provided by the metaphoric topic and the linguistic context prior to the construction of larger meaning structures.

Despite the large body of experimental data collected on this issue, it has not been possible to settle the debate and tip the scale in favor of one or the other set of theories (see Pouscoulous and Dulcinati (2019) and Holyoak and Stamenković (2018) for comprehensive reviews). So how do listeners perform the feat of understanding a novel metaphor?

## 1.1 Motivation

The central idea of this dissertation is that the debate on metaphor processing can benefit from an empirical investigation of previously understudied metaphors, namely **non-nominal metaphors**, that takes into consideration the role that context plays during processing and incorporates the findings and methodology of the field of situated and incremental language processing. The motivation is to take a new perspective in order to move the debate on metaphor processing forward.

One issue with the debate on metaphor processing is that, by large, it has been centered around nominal metaphors of the canonical form “X is a Y”, as in sentence (1), which are also almost always studied in the absence of a linguistic context. On the other hand, very little to no attention has been payed to the processing mechanisms at play in metaphoric expressions such as (2b) or (3), in which topic and vehicle are not both nouns, and in which the metaphoric interpretation is not brought on by the interaction between topic and vehicle alone but by the interaction between context, topic and vehicle (as is the case of 2b).

Nominal metaphors, though easy to create and to test in the lab, impose restrictions on the type of hypotheses that can be tested. For one thing, the syntax of a nominal metaphor is suggestive of Category Inclusion (“X is a Y”). Furthermore, it is not possible to reverse the sequential order of presentation of the elements of a metaphor (topic and vehicle) without changing the meaning of the expression as a whole. This gives researchers limited leeway when it comes to investigating what the elements of the metaphor individually contribute to the comprehension process and when.

It is also important to note that language comprehension does not happen in a vacuum: It is instead embedded in and contingent upon a situational and conversational context that shapes its interpretation. Studying nominal metaphors outside of a context thus brings with it a lack of ecological validity that limits the scope of theory development. In addition to this limitation, investigations on metaphor processing typically involve the measuring of reaction times once the entire metaphoric expression has been understood. This coarse measure of processing time does not allow us to test theoretical predictions that require a more time sensitive measure.

One issue for which these limitations are visible is that of **metaphoric symmetry**: The Implicit Comparison view claims that both elements in a metaphor are processed equally and thus their relative sequential order should not matter, making the understanding of the metaphor a symmetric process (at least during initial stages of processing). The Category Inclusion view, on the other hand, sees metaphor comprehension as an asymmetric process and claims that both elements play different roles from very early on during sentence comprehension, making their relative position essential for understanding. However, vehicle and topic of nominal metaphors cannot be reversed without rendering the entire expression infelicitous or changing its meaning altogether (*some lawyers are sharks* is not equivalent to *some sharks are lawyers*). Furthermore, previous investigations on this matter have not considered the way in which context interacts with the relative position of the metaphoric elements as these unfold over time. As a result, the theories remain underspecified in this regard.

To resolve this, it would be necessary to examine the processing of contextualized metaphors that do allow for a felicitous reversal of the order of its elements. In addition to this, given that the theoretical claims refer to the individual contribution of the elements, it should be studied how metaphoric processing unfolds over time with adequately time-sensitive technology.

Another issue where the limitations of previous studies are hindering the resolution of the theoretical debate concerns the **role of literal features of a metaphor** during comprehension. Experiments testing between theories of metaphor processing have mostly been carried out using priming paradigms where both prime and target are sentences or words and reading or reaction times are measured (Gernsbacher et al., 2001; McGlone & Manfredi, 2001; Rubio Fernandez, 2007; Keysar et al., 2000; Thibodeau & Durgin, 2011; but see Weiland et al., 2014, for an ERP version of this design). These paradigms have lead to contradictory claims about what happens to features associated with the literal meaning of a metaphoric vehicle during processing. Processing theories make different predictions in this matter: Category Inclusion views hold that features of a metaphorically used word that are relevant for the literal meaning but irrelevant for the metaphorical meaning - “has fins”, or “lives under water” in sentence (1), for example - should be suppressed during comprehension and not remain active once the metaphor has been understood (Glucksberg, 2001, p. 66). Implicit Comparison views, on the other hand, believe that all features associated with the encoded meaning of a word have to be activated in order for structural alignment to take place. This includes those features that are associated only with the literal meaning and are irrelevant for the metaphoric interpretation. These literal features are only suppressed in secondary comprehension phases once structural alignment has been completed (Wolff & Gentner, 2011).

Though some studies show that irrelevant features are rapidly suppressed during processing (Rubio Fernandez, 2007) and hinder subsequent metaphoric comprehension (McGlone & Manfredi, 2001) others have shown that they can in fact facilitate subsequent metaphoric comprehension (Weiland et al., 2014) and even remain active after comprehension has taken place (Thibodeau & Durgin, 2008). How do we overcome this impasse? One possibility would be to examine this question by looking at a previously unexplored type of metaphor, such as verbal metaphors like the one in sentence (3) above. This would allow us to see which of the previous findings on nominal metaphors can be generalized to non-nominal metaphors, which would in turn strengthen the respective theoretical position.

A third unresolved issue worth examining is the role of **conventionality, familiarity and aptness** of a metaphoric expression during comprehension. **Metaphoric aptness** is defined as the degree to which the figurative meaning of a metaphoric vehicle captures relevant features of the metaphoric topic (Jones & Estes, 2006). **Metaphor conventionality** is determined by the frequency in which a metaphoric vehicle is used in its metaphoric meaning in a given language (Bowdle & Gentner, 2005), while **metaphoric familiarity** captures the frequency of use of the entire metaphoric expression, i.e. of topic and vehicle (Thibodeau & Durgin, 2011).

As shown by Jones and Estes (2006), several studies on metaphor processing claim that both conventionality and aptness facilitate the processing of nominal metaphors (Blasko & Connine, 1993; Bowdle & Gentner, 2005; D. Chiappe & Kennedy, 1999; D. Chiappe, Kennedy, & Smykowski, 2003; Jones & Estes, 2005). However, when taking a closer look at these studies, aptness is consistently correlated with conventionality, making it difficult to determine which of the two mediates metaphor processing best. This is a critical point for theory development given that Category Inclusion views see aptness as the true mediator of processing mode (Glucksberg, 2008, p. 80; Jones & Estes, 2006, p. 19): More apt metaphors will be easily processed via category inclusion whereas less apt metaphors might be processed via analogical reasoning. Implicit Comparison views, on the other hand, attribute this mediating role of processing mode to conventionality (Bowdle & Gentner, 2005, p. 199): Highly conventional metaphors will be processed via category inclusion and less conventional metaphors via analogical reasoning. Additionally, Thibodeau and Durgin (2011) suggested that not conventionality nor aptness modulate processing mode, but familiarity, with more familiar metaphors being processed via category inclusion and less familiar ones via analogical reasoning.

The effect of these three variables, however, has yet to be systematically studied within the same data set. Furthermore, all evidence in this debate has come exclusively from de-contextualized nominal metaphors. Taking this into consideration, a path forward towards resolving this issue could be to examine how the three variables systematically affect processing of different types of non-nominal metaphors, such as examples (2b) and (3). Furthermore, this presents an opportunity to refine existing theories of metaphor comprehension by accounting for the role of the linguistic context during processing.

Overall, this dissertation aims to make a contribution to the field of metaphor processing by re-examining these three issues in a way in which it has not been been previously attempted, namely by investigating the processing of non-nominal metaphors, which have been largely ignored in previous research: Verbal metaphors, such as in sentence (3), and verb-object metaphors, such as in sentence (2b). I investigated their processing using the eye-tracking method in two eye-tracking during reading studies (Rayner, 2009), in which participants' gaze patterns are recorded while they read at their own pace, and two Visual World studies (Cooper, 1974; Tanenhaus et al., 1995), in which participants' gaze patterns are recorded while they hear spoken language and simultaneously view a grid of four images.

These studies are complemented by two experiments measuring reaction times, six offline rating tasks, and two re-analyses of the previously collected eye-tracking data. Taken together, these 14 empirical studies aim to shed light on the mechanisms involved in the process of understanding how metaphoric meaning is constructed during language comprehension. The studies were conducted on two sets of novel experimental materials that were tailored to address the three theoretical issues previously mentioned: The issue of symmetry, the role of literal features and the differential contribution of aptness, conventionality and familiarity during processing. Additionally, these studies allow for an investigation of the role of context during metaphor processing, a topic which has often been neglected for the purpose of theory development.

More specifically, the studies in this dissertation aim to test the predictions made by the two theoretical positions introduced earlier. The position of each set of views on the three main topics are broadly summarized in Table 1.1.

**Table 1.1:** Differences of the two theoretical views on the three main issues of this dissertation

| **Empirical issue** | **Position of Category Inclusion View** | **Position of Implicit Comparison View** |
|---|---|---|
| *Role of literal features* | Suppressed at earliest stages of processing. Should cause interference if pre-activated | Suppressed only in later stages of processing. Should facilitate (early stages of) processing if pre-activated |
| *Symmetry of processing* | Processing is asymmetric. reversal of position of the elements should cause difference in earliest stages of processing | Processing is symmetric. Reversal of position of the elements should not cause differences in earliest stages of processing. |
| *Mediating factors* | Aptness is true mediator of processing mode. | conventionality or familiarity mediate processing mode. |

The issues are addressed by incorporating insights from the literature on situated language processing, which has shown that visual scenes can rapidly influence language comprehension during early stages of processing, even during processing of abstract language as well as complex pragmatic phenomena (see Knoeferle & Guerra, 2016, for a review). I also incorporate the insights of the literature on incremental language processing (e.g. Altmann & Steedman, 1988; Eberhard et al., 1995; Van Berkum et al., 2005; Kamide et al., 2003; Kamide, 2008; Tanenhaus et al., 1996), which have shown that language comprehenders rapidly integrate incoming linguistic input with their available linguistic and non-linguistic context in an incremental way. The goal of assuming both a situated and an incremental language processing perspective is twofold: First, it might help us resolve the three open issues displayed in Table 1.1 by introducing new technologies and paradigms into the theoretical debate on metaphor processing[1]. Second, it might help us better understand the role that context plays during metaphor processing

[1]This is not to say that highly time-sensitive technology has never been used before to study metaphors (e.g. Bambini et al., 2016; Coulson, 2012; Coulson & Lai, 2016; Lai et al., 2009; Lai & Desai, 2016; Pynte et al., 1996). It has not, however, been used previously to test these theories on these specific issues, with the exception of Weiland et al. (2014).

by comparing it to the way in in which context has been shown to affect processing of other linguistic phenomena.

## 1.2 Thesis Outline

The thesis is structured as follows: Chapter two presents the development and the state of the art of the current debate on metaphor processing, focusing on the Implicit Comparison view and on the Category Inclusion view. The relevant empirical studies are discussed with a focus on the three critical issues shown in Table 1.1. I discuss the limitations of the way in which these two theoretical views have been empirically tested, as well as the way in which the theories address the role of context during processing. This chapter also introduces the benefits of incorporating the perspective of situated and incremental language processing into the study of metaphor processing in order to move the theoretical debate forward.

Chapter three presents the first set of studies, which address the issue of the role of features of the literal meaning during comprehension of novel verbal metaphors. Specifically, the studies examine whether the concept of physical containment is activated during and immediately after the processing of verbs of physical containment that are used metaphorically to signify difficulty, such as *fenced in* in sentence (3).

Chapter four presents the second set of studies, which investigate the time-course of comprehension of novel verb-object metaphors and the role of the linguistic context during processing. These studies directly address the issue of symmetry: Are metaphors understood differently as a function of the relative position of their elements? To answer this question I draw from the insights of the literature on incremental language processing and make use of the Visual World Paradigm.

Chapter five investigates the influence of aptness, familiarity and conventionality on the processing of non-nominal metaphors. Two experiments from chapters three and four are reanalyzed incorporating the results from six novel experiments aimed at teasing apart the effect of these three variables on the processing of novel metaphors.

Finally, chapter six presents the conclusions drawn from the individual chapters and evaluates them in the light of the debate between theories on metaphor processing. This chapter also presents a tentative outline of the way in which existing theories can be refined to account for the results of this dissertation.

# 2
# Accounting for Metaphor Comprehension

This chapter presents the two leading accounts on metaphor processing at the center of this dissertation: The Implicit Comparison View and the Category Inclusion View. The goal is to critically examine how these views relate to the three core issues of this dissertation, presented in Table 1.1 of chapter one: The issue of symmetry, the role of literal features during processing, and the effect of mediating factors.

However, before diving into these issues, it is crucial to acknowledge the historical context that brought about both sets of theories being discussed. Before the second half of the twentieth century, the study of metaphors was a part of the study of literature and rhetoric, and not of linguistics or psychology (see Gibbs, 1993). This changed with the work of the philosopher Paul Grice (Grice, 1989), who attempted to account for the way in which people are able to understand metaphors, irony, and other types of figurative language. After this, a lot of the subsequent work became concerned with empirically testing (and eventually refuting) the psychological plausibility of the Gricean view on figurative language comprehension, in what came to be known as the "psycholinguistic turn" (Pouscoulous & Dulcinati, 2019, p. 315). It was as a response to the Gricean views, then, that both the Implicit Comparison View and the Category Inclusion View came to be.

Because of the important role played by the Gricean program in the historical development of theories of metaphor comprehension, this chapter begins with a discussion of Grice's ideas. I will then discuss the Implicit Comparison View and the Category Inclusion View individually, with a focus on the way in which each view accounts for the role of context during metaphor processing. This will be followed by a critical examination of these two views, assessing their predictions on the three theoretical issues displayed in Table 1.1. Finally, I will attempt to show how a new approach could help to both resolve these issues as well as to further refine the theories: An approach that examines the processing of non-nominal metaphors from the perspective of situated and incremental language processing. This will lay the groundwork for the remaining chapters of this dissertation.

## 2.1 The Standard Pragmatic Model

### 2.1.1 Paul Grice's account of figurative language understanding

Prior to the mid-twentieth century, metaphors were seen as just one phenomenon among many on a list of rhetorical devices that could be used to make a written text more aesthetically pleasing or rhetorically convincing. It was understood as self-explanatory that metaphors represented a transfer of meaning triggered by comparing two terms, a view traditionally attributed to Aristotle, who described metaphors this way in both *The Poetics* and in *Rhetoric* (as cited by Kirby, 1997; and Levin, 1982). Scholars of figurative language were more concerned with categorizing the different instances of this transfer-of-meaning process than with the questions of why and how we understand metaphors.

This changed with the work of Grice (Grice, 1989), who attempted to account for figurative language understanding by incorporating it in his comprehensive account of communication. Grice's analysis of language use emerges from the observation that language is a cooperative activity. During conversation, speakers plan and organize their conversational contributions in order to achieve a common goal with their listeners. By doing this, they generate expectations in the minds of their listeners, given that they share these communicative goals and assume that their conversational partners will act accordingly. Grice formulated this idea as the Cooperative Principle of communication ("Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged") (Grice, 1989, p. 26), and further explicated it by stating the precise type of expectations that interlocutors generate during conversation, known as the Conversational Maxims (Grice, 1989, p. 26-27):

- **Maxims of Quantity** Do not make your contribution more or less informative than is required.
- **Maxims of Quality** Make your contribution one that you believe to be true (do not say what you believe to be false).
- **Maxim of Relation** Make your contribution one that is relevant to the conversation.
- **Maxims of Manner** Be perspicuous, avoid obscure expressions and ambiguity.

The four maxims and the cooperative principle are not to be taken as deterministic laws of language. They are a set of reasonable assumptions made by conversational partners that can account for the systematic differences between the linguistic meaning of sentences (i.e. the conventional meaning of words plus the grammatical rules that combine them) and the meaning a speaker intends these sentences to take on during conversation (what Grice called "Speaker's meaning"). For a listener, grasping a speaker's meaning is a process of inference to the best explanation on the basis of the Cooperative Principle and the conversational Maxims: When the listener perceives that the speaker is not observing the maxims with a given utterance, it is reasonable to infer that the speaker meant to convey additional information that can be inferred during conversation from said utterance. Grice called these additional pieces of meaning **implicatures**.

Under this model, metaphoric expressions fall under the type of implicatures that depend on a listeners's ability to recognize that a speaker has overtly violated the Maxim of Quality (Be truthful!) (Grice, 1989, p. 34): If a speaker utters something that is literally false, the listener must draw the inference that the speaker actually meant to communicate a different proposition. So, if the speaker utters *My lawyer is a shark*, the listener, after determining the falsehood of the literal proposition, infers that the speaker meant to attribute shark-like qualities to his lawyer via transfer of meaning (under the assumption that the speaker is being cooperative). This leaves us with a theory of metaphor in which a listener first computes the truth-conditional meaning of an utterance, evaluates it given the state of the world and subsequently reinterprets it.

### 2.1.2 Cognitive architecture of the Gricean account

The Gricean account of metaphor comprehension was never meant as a real-time psychological description. The goal was to provide a plausible reconstruction of the necessary logical steps that a listener undertakes when going from an utterance's meaning to a speaker's meaning (Bach, 2006). However, the Gricean account did lay the foundations for what later came to be known as **The Standard Pragmatic Model (SPM)** (Gibbs, 1986), which is a translation, in psychological terms, of the Gricean account together with John Searle's ideas on metaphor comprehension (Searle, 1979). Searle's basic tenet is that metaphors (and other forms of figurative speech) are a type of indirect language use that is understood only if a literal interpretation of an utterance is considered "defective" and must be discarded (Searle, 1979, p. 114). In psychological terms, understanding metaphors would then have to be a three-stage process: The first step requires the computation of the literal meaning of the utterance (i.e. the truth-conditional output of assembling the conventional meaning of the individual words with the grammatical combinatorial rules). In the second stage, the output of stage one is evaluated and discarded. In the third stage, the utterance is re-evaluated given context and the appropriate implicature is derived. Each of these stages must be seen as requiring a specific amount of cognitive effort, operationalized as processing time, which results in the conclusion that deriving metaphoric meaning (a three-stage process) would necessarily take longer than understanding an equivalent literal utterance, which would only need the first step of evaluating the truth of the proposition (Gibbs, 1979, 1984).

### 2.1.3 Arguments against the SPM

The SPM represented a step forward in figurative language research given that, for the first time, metaphor is analyzed on par with everyday speech and not seen as a rhetorical device used only by poets. The SPM also provided a novel explanation of why we are able to understand and use metaphors (and figurative language in general) in a reliable way. However, the account has been extensively criticized both from a theoretical (Carston, 2012; Gibbs, 1984; Recanati, 2004; Wilson & Sperber, 2012, i.a.) and en empirical (Gibbs, 1986; Glucksberg et al., 1982; Ivanko & Pexman, 2003, i.a.) perspective, which ultimately lead researchers to reject the SPM for metaphor processing and processing of figurative language in general.

The theoretical criticism boils down to the SPM suffering from the same issues that apply to the traditional Aristotelian view on figurative language. First of all, the SPM does not provide an explanation as to why a speaker would choose to use a metaphoric expression instead of a literal one given that a metaphorical utterance is assumed to require considerably more effort on the part of the listener to comprehend. This would seem to be at odds with the goal of the Gricean program, which intends to defend the generating of implicatures on a rational basis.

A more specific problem of the Gricean account regarding metaphor comprehension is the lack of explanatory power that comes from an appeal to the Maxim of Quality. If the main mechanism at play during figurative language comprehension is the recognition of a sentence being literally false and re-analyzing it in context, what is the difference, from a cognitive perspective, between understanding metaphors and other types of figurative language, such as irony? How does a language user know which figurative meaning to derive? The SPM does not provide an answer to this question.

Furthermore, the Gricean account of metaphor comprehension is also inconsistent with the rest of the Gricean communicative agenda. Wilson and Carston (2006) points out that according to Grice, deriving speaker-meaning from an utterance is an additive operation: Listeners make an inference given the sentence-meaning and reinforce it with the stronger, implicated meaning. In (1) below, The weaker meaning of some (some and possibly all) is reinforced with a stronger one (some and not all) after computing what is known as a scalar implicature:

*1) Carlos asked Maria if she had eaten all of his Nutella. Maria said she ate some of it.*

In the case of metaphor, we can't speak of an additive process, but of one where the sentence-meaning is in some way replaced by the derived speaker-meaning. In (2) below, it is not the case that the literal meaning of Charlie being a large book with legs is extended by the inference that Charlie is also very knowledgeable. The operation here would have to be one of replacing one meaning for the other, with no explanation as to why metaphors should behave so differently than other types of expressions with inferential meaning.

*2) Charlie is really a walking encyclopedia!*

Asides from theoretical concerns, the main reason why the SPM is no longer considered to be a viable cognitive model of metaphor comprehension is the weight of the empirical evidence against it. The main empirical finding in this regard is that understanding metaphors and other types of figurative expressions does not necessarily take longer than understanding equivalent literal utterances. Ortony et al. (1978) had participants read metaphoric and literal sentences embedded in either short or long supportive contexts. They found that though there was a reading-times advantage for literal sentences when following the short contexts, this advantage disappeared in the long context conditions. This is incompatible with the Gricean account that posits an automatic initial processing of the literal meaning of every figurative sentence, regardless of its compatibility with context. In a study on verbal irony, Gibbs (1986) reported a similar finding: When embedded in a supportive context, sentences were read just as fast regardless of whether they were intended literally or ironically.

A crucial finding that represents the nail in the coffin for the SPM is that by Glucksberg et al. (1982) on the automaticity of metaphor processing. Inspired by the Stroop task, they had participants read metaphoric, literal and nonsensical statements. Participants were instructed to answer whether the sentences were literally true or false by pressing either a YES or NO button in front of them. The logic of the study was as follows: According to the SPM, computing the truth conditions of an utterance is an obligatory first step in processing, with the derivation of a metaphorical implicature being an optional secondary step. If this is the case, it should be equally easy for participants to recognize that a nonsensical sentence such as *Some jobs are birds* is false compared to the metaphor *some jobs are jails*: In both cases, the truth conditions of the literal meaning should be computed first, so both should be easily recognized as literally false. However, Experiment 1 of Glucksberg et al. (1982) showed that this is not the case. Participant took significantly longer to identify metaphors as false compared to nonsensical sentences, suggesting that participants were automatically processing the metaphorical meaning, causing a response delay. The authors argued that the results speak for the automatic, obligatory character of metaphoric meaning derivation and against the SPM's three-stage process.

The SPM started the conversation on the mechanisms involved during metaphor comprehension. Since it was proven to be inadequate on both theoretical and experimental grounds, it became necessary to look elsewhere for models that are both cognitively plausible and supported by empirical evidence. There are currently two sets of views that fulfill these criteria (Holyoak & Stamenković, 2018): The metaphor-as-category-inclusion and the metaphor-as-implicit-comparison views. In what follows I will describe each of these views. Following this, I will evaluate the evidence on the three central points of this dissertation: The role of the literal meaning during processing, the symmetry of metaphors, and the mediating effect of aptness, conventionality and familiarity.

## 2.2 Metaphor as Category Inclusion

The basic tenant of Category Inclusion models is that nominal metaphors are processed just as their syntactic form would suggest: as category statements, in which the first element (the metaphoric topic) is understood as a member of the category represented by the second element (the metaphoric vehicle). This idea was first present in the work of Roger Brown (Brown, 1958, p. 140), who discussed how the word *foot* can be extended from its original meaning (*a person's foot*) to a super-ordinate category (*the foundations or lower parts of things*) that encompass both the original base-level meaning and a metaphorical one such as in the expression *the foot of the mountain*. Brown's ideas were only fledged out decades later, and there are currently two accounts based on this premise[1]: Sam Glucksberg's Dual Reference model (Glucksberg, 2001) and Sperber and Wilson's Deflationary Account of metaphor comprehension (Sperber & Wilson, 2008), which is embedded in the larger cognitive framework of Relevance Theory (Sperber & Wilson, 1986; Wilson & Sperber, 2012).

### 2.2.1 Dual reference

Following their groundbreaking finding of the automaticity of metaphor processing, Glucksberg and colleagues set out to develop a novel account of metaphor comprehen-

[1] Others have postulated similar accounts on metaphor comprehension, such as François Recanati (Recanati, 2004). However, Glucksberg's and Sperber and Wilson's accounts are the most explicit in terms of the mental processes and representations involved, generating empirically testable predictions.

sion that was compatible with this empirical finding as well as with other findings and claims made throughout the late 1970's and 1980's stating that processing figurative language was not necessarily more difficult than processing literal statements (Gibbs, 1986; Lakoff & Johnson, 2008; Ortony et al., 1978). This resulted in what came to be known as the "dual reference" or "class inclusion" account (Glucksberg, 1989, 2001, 2008; Glucksberg & Keysar, 1990; Glucksberg et al., 1997).

The starting point was the rejection of the primacy of literal meaning: Understanding metaphors cannot be a process of first computing the literal meaning, discarding it and reaching a metaphoric interpretation via indirect inference. Instead, these authors accounted for metaphoric meaning by appealing to a process of **dual reference**: During processing, the meaning of a metaphoric vehicle is spontaneously altered in order to refer to a category with a higher order level of abstraction of which the metaphoric vehicle is a prototypical member. The expression as a whole is understood as a category statement, indicating that the metaphoric topic is a member of the category represented by the vehicle.

There are many established cases of words prototypically standing in for a higher-level category in languages across the world (Newport & Bellugi, 1978). For example, in American Sign Language (ASL), categories are expressed by referring to their most common member.

*3) HOUSE FIRE ['t-] LOSE ALL CHAIR-TABLE-BED, ETC., BUT ONE LEFT, BED*

*"I lost all my furniture in the house fire but one thing was left: the bed"*

Newport and Bellugi (1978) used example (3) above to show that the ASL signs for "chair", "table" and "bed" can stand in for the "furniture" category as a whole. This same phenomenon can also be seen to have occurred in Hebrew, where the word *Demjanjuk*, meaning "an ordinary person capable of committing unspeakable acts" was coined after the trial of war criminal John Demjanjuk, accused of having been a guard at the Treblinka death camp in Poland during the Second World War (Glucksberg & Keysar, 1990).

Less somber examples can easily be found in the genericization of brand names, i.e. the process through which the name of a specific brand comes to represent all products of the same category. In German, for example, the brand name *Tempo* is used to refer to the entire category of disposable tissues. The brand name *Labello* refers to the entire category of lip balms, which in English can also be called *chapstick* and in Swedish *Lypsyl*, which are also brand names. This pattern of meaning extension is not restricted to brand names, as illustrated by one poignant example in Glucksberg et al. (1997):

*4) Cambodia was Vietnam's Vietnam*

The first use of the word *Vietnam* in (4) refers to the base-level meaning (i.e. the country), whereas the second refers to the superordinate category of "failed, dragged-out military interventions". Despite this particular novel use of the word *Vietnam*, someone acquainted with the relevant political and historical context should nevertheless be able to easily understand the meaning of (4) as a whole.

An important property of (4) that makes it different from the genericization examples of *Tempo* and *chapstick* is that the intended superordinate category in (4) is constructed on the fly and is not yet lexicalized. Glucksberg and Keysar (1990) claim that metaphors instantiate precisely this type of **ad hoc categories**. Independently of metaphoric expressions, it has been noted that people often make use of newly-created categories to serve a specific purpose. For example, the category of "things that are necessary during a pandemic" could include concrete objects such as hand sanitizers and surgical masks as well as more abstract ones such as the idea of social distancing. This category might not be lexicalized (yet), but it can prove to be quite useful to understand and maintain during the event of an actual wide-spread infectious airborne disease.

Barsalou (1983) showed that ad hoc categories such as "things you can bring to a picnic" - which are less established in memory than lexicalized categories - can nevertheless be processed similarly to regular lexicalized categories in the way that they display a coherent internal structure. Barsalou (1983) had participants rate different potential members of both lexicalized and ad hoc categories. The results showed that, for both types of categories, participants recognized some members as being more typical than others, a property of categories known as gradient typicality.

Inspired by Barsalou (1983)'s findings, the suggestion made by Glucksberg and Keysar (1990), explicated in, Glucksberg (2001, p. 52-67), is that, in a nominal metaphor such as *My job is a jail*, the word *jail* represents both the abstract, ad hoc, category of "involuntary, unpleasant, confining, punishing, stifling situations" (which has the properties of ad hoc categories described by Barsalou, 1983) and the base-level member of the category *prison*, which is a prototypical instance of the superordinate ad hoc category. The ad hoc category is referred to as JAIL* .

But how does a comprehender know exactly which ad hoc category to create? The word *snake* for example, could be used metaphorically to describe either people who are vicious in character and potentially dangerous or something that is shaped the way a snake is.

*5) My lawyer was a snake.*

*6) The road was a snake.*

Sentences (5) and (6) (taken from Glucksberg, 2001, p. 53) show just that. The ad hoc category is in each case a different one, and one could not say that the *snake* in (5) counts as a member of the category described by *snake* in (6). Glucksberg (2001) (p. 53) claims that in order to create the correct ad hoc category in each case it is necessary to use a **set of dimensions for attribution** provided by the most meaningful and relevant features of the metaphoric topic as parameters that regulate the modulation in meaning of the metaphoric vehicle. This means that the meaning of the metaphoric expression is derived from an interaction between topic and vehicle in which each of them plays a different role: The topic provides dimensions for attribution and the vehicle changes its meaning accordingly. This is described as a process of interactive property attribution by Glucksberg et al. (1997).

### 2.2.2 The role of context in the formation of ad hoc categories

One problem with the dual reference account is its apparent lack of consideration of the role that context plays during comprehension. This is illustrated in examples (7) and (8) below.

*7) My brother always eats more than he should and now he is very overweight. My brother is an elephant.*

*8) My brother can remember anything you tell him because of his incredibly good memory. My brother is an elephant.*

(7) and (8) have identical metaphoric topics (*My brother*) and metaphoric vehicles (*elephant*). However, they have very different metaphoric meanings because they exploit different salient properties of elephants: Being very fat (7) or having a good memory (8). Glucksberg (2001, p. 54) claims that this is due to the fact that some topics have more relevant dimensions of attribution that others. *Lawyer*, for example, is likely to be characterized on few dimensions (e.g. skill, reputation, experience, ambition, cost) and it therefore imposes strict constraints on how the ad hoc category might be constructed, whereas *brother* provides very few constraints (one could say a lot of different things about one's brother), forcing the construction of the ad hoc category to depend on contextual information.

However, if the word *brother* is replaced with the word *lawyer* in examples (7) and (8) the problem remains, as illustrated in examples (9a) and (9b):

*9a) My lawyer always eats more than he should and now he is very overweight. My lawyer is an elephant.*

*9b) My lawyer can remember anything you tell him because of his incredibly good memory. My lawyer is an elephant.*

Here, we have two identical topic-vehicle pairings with very different interpretations, regardless of how narrow the set of constraints that the word *lawyer* is believed to impose. As a result, it is necessary for the theory to consider the role that context plays in the modulation of the meaning of the metaphoric vehicle.

Context doesn't just play a role on meaning modulation of novel metaphors, but has been shown to also modulate sense selection of ambiguous words. Tabossi (1988) and Tabossi and Zardon (1993), for example, showed that initial access of an ambiguous word can be influenced by prior context when the context makes salient a characteristic feature of a word's dominant meaning: Contextually priming the dominant meaning of the word *port* (a harbor) by mentioning the word *ship* in the context facilitates lexical access to the word *sea* in a subsequent lexical decision task more so than to the word *liquor*, which is related to the subordinate meaning of *port* (a type of wine) or to an unrelated control word (*hand*) (Experiment 3 of Tabossi, 1988). There was also no difference in priming effect between subordinate (*liquor*) and control (*hand*) words. However, deprived of a context or in the presence of a neutral context that did not bias the interpretation of the target word, both dominant (*sea*) and subordinate (*liquor*) words displayed priming effects relative to the control word (*hand*) (Experiments 1 and 2), in line with previous findings (Swinney, 1979).

Context-sensitivity of conceptual features has also been found in instances of conceptual combination. Glucksberg and Estes (2000) investigated whether conceptual combinations such as *peeled apple* would make features of the head noun (*round*, i.e. the shape of apples) or features of the adjective-noun phrase (*white*, i.e. the color of peeled apples ) more accessible as a function of said feature's contextual relevance. They set up an experiment in which participants read a 4-sentence context that made either noun-level (some kids are looking for balls to play bowling with and chose to use peeled apples) or phrase-level properties (a chef is making a colorful fruit arrangement and uses peeled apples) relevant (context factor: noun-relevant vs. phrase-relevant). Participants then verified statements about either phrase-level (*peeled apples are white*) or noun-level (*peeled apples are round*) features (feature factor: phrase vs. noun). They found that in the noun-relevant context conditions, participants were faster at verifying noun compared to phrase features. The reverse was true in the phrase-relevant context conditions: Participants were faster to verify phrase features compared to noun features. This resulted in a significant interaction between the two factors, but in no main effects. The authors interpreted this result as suggesting that during processing of conceptually combined phrases such as *peeled apples*, relevant information is more accessible than irrelevant information, independently of whether these features relate to the level of a phrase or of the phrase's head noun.

Metaphoric expressions could also be seen as a type of contextually sensitive conceptual combination: In (9a) and (9b) the words *lawyer* and *elephant* are combined to produce a new interpretation of the expression as a whole. In each case, however, the interpretation heavily depends on the properties of *lawyer* that are made relevant by the context and not by any inherent set of dimensions provided by the word *lawyer* on its own: "Fat" in (9a) and "good memory" in (9b). Accordingly, it can be said that contextual relevance affects the availability of salient features of the metaphoric topic, which in turn provides a different set of parameters necessary to modulate the meaning of the metaphoric vehicle, resulting in the two very different interpretations of (9a) and (9b).

The role that context plays in restricting the set of parameters for modulation is more thoroughly explicated by another account that views metaphor comprehension as a type of category inclusion: Sperber and Wilson's Deflationary Account of metaphor comprehension (Sperber & Wilson, 2008).

### 2.2.3 Sperber and Wilson's deflationary account

Sperber and Wilson (2008) lay out an account that is both similar to and compatible with the dual reference model of Glucksberg and collaborators. They posit that metaphors are understood through a process of **lexical modulation**: The vehicle in a nominal metaphor undergoes adjustment of its lexical meaning given a set of contextual parameters. This process results in the creation of an occasion-specific, ad hoc category, very much in line with Glucksberg's view. There are, however, two key aspects in which the views differ, namely the scope of Sperber and Wilson's account and the degree of specification of their theory.

#### Scope of the account

Sperber and Wilson claim that metaphors are not a natural class, but are instead on a continuum of cases in which lexical meaning is contextually adjusted (thus the use of the word "deflationary" when referring to the account). This means that Sperber and Wilson's account is considerably broader in scope than Glucksberg's, since Sperber and Wilson believe that constructing ad hoc concepts is not only the way through which metaphors are understood, but more generally how most words are pragmatically enriched. Ultimately, the goal of this account is to unify all processes involved in lexical pragmatics (Wilson & Carston, 2007). To do this, they make use of the inferential machinery of the communicative framework known as **Relevance Theory** (Sperber & Wilson, 1986; Wilson & Sperber, 2004; Wilson & Sperber, 1999; Wilson, 2003; Wilson & Carston, 2007; Wilson & Sperber, 2012, i.a.).

Relevance Theory takes two Gricean ideas as a point of departure, namely that (1) a fundamental feature of human communication is the expression and understanding of intentions and (2) that in conversation, utterances create expectations that make it possible for the listener to understand what a speaker's intentions are (Wilson & Sperber, 2004). As explained in the beginning of this chapter, Grice believed that a listener understands a speaker's intentions by reasoning about the speaker's utterance on the basis of the cooperative principle and the maxims of conversation. Sperber and Wilson overhaul this system by claiming that listeners only need a general principle of relevance to guide their inferential reasoning (Sperber & Wilson, 1986, p. 46).

According to this Principle of Relevance[2], every utterance brings with it an expectation of optimal relevance that provides sufficient evidence of the underlying communicative intentions of a speaker. Relevance is defined as a property of utterances (and other stimuli) to cognitive processes: An utterance is relevant when it minimizes the cognitive effort necessary to process a given stimulus and maximizes the cognitive effect achieved by processing said stimulus. The greater the effect and the smaller the effort, the more relevant an utterance is to a hearer. For our discussion on metaphor, the most important type of cognitive effect is that of a contextual implication (see Wilson & Carston, 2007), which is derivable by integrating an utterance with the context available to the listener.

The relevance-oriented comprehension heuristic is what allows a listener to understand the meaning of *asleep* in sentence (10) below.

*10) The audience was asleep throughout the professor's lecture.*

In (10), the speaker might mean to communicate that all members of the audience literally had their eyes closed during the lecture and slept throughout it (literal, conventional meaning of *asleep*). It is more likely however, that the intention was to communicate that only some of them were sleeping or perhaps yawning (ASLEEP*), in which case the speaker might be exaggerating, or that the audience was merely bored or not attentive (ASLEEP**), in which case the

[2] The Principle of Relevance comes in two versions: the Communicative Principle, which refers specifically to the way in which utterances carry the expectation to be relevant enough to be processed by the listener, and the Cognitive Principle, which refers to the way in which cognition in general is geared towards optimal relevance. Only the Communicative Principle is of interest for the current discussion.

speaker was speaking metaphorically. The correct interpretation depends on the set of interpretative assumptions available to the listener: If, for example, the listener knows that the professor is famous for his tedious lectures, she might more readily understand *asleep* as standing in for the ad hoc category ASLEEP**, meaning something along the lines of "boring", because this interpretation is the most relevant one given her assumptions: It strikes an optimal balance between the cognitive effects it brings with it and the cost of deriving them.

As the explanation of example (10) shows, Relevance Theory provides a specific account regarding the way in which context is integrated during utterance comprehension: To determine the appropriate meaning extension of *asleep*, the listener must incorporate their knowledge about the speaker's intentions, which in turn are inferred from the situational context. This process is triggered by every utterance's presumption of optimal relevance. For metaphor processing this means that the modulation of a vehicle will always be contingent on the relevant context and is not only a product of the topic-vehicle interaction.

### Degree of specification

Sperber and Wilson's view is more specific than Glucksberg's in that it provides a more detailed classification of the different types of lexical modulation as well as a more systematic view of the role of context during comprehension. According to Sperber and Wilson (2008), there are two distinct types of lexical modulation involved in metaphor comprehension (and in lexical pragmatics in general): Lexical broadening and narrowing. Through broadening, lexical meaning is extended so that the newly-constructed ad hoc concept expresses a broader set of senses than the encoded meaning would. Through narrowing, the ad hoc concept becomes more specific, excluding cases that the encoded meaning would cover.

*11) It's 2 PM right now.[uttered at 1:56 PM]*

*12) I'm freezing. [uttered at 20 degrees Celsius]*

*13) Psycholinguists like to drink. [uttered at a post-conference social event]*

In (11) we have a case in which a number has an imprecise, approximate interpretation: the encoded meaning is broadened to accept numbers that are close to 2 PM if it meets a listener's expectations of relevance. In (12) we have a more extreme case of broadening: The word *freezing* changes meaning to include a wider range of temperatures. In (13) we have a case of lexical narrowing: *Drink* no longer refers to ingesting liquids in general, but specifically to drinking alcohol, and perhaps also plenty of it.

There is empirical evidence supporting the view that speakers modulate the meaning of words when uttering sentences such as (11) by considering the listener's expectations of relevance (i.e., by inferring the interlocutor's intentions through the linguistic and situational context). Gibbs and Bryant (2008) investigated the way in which people produce answers to questions about time. They found that when the question was framed without a context (e.g. *Could you tell me what time it is?*) participants tended to round their answers, as in sentence (11). However, when speakers were given different clues regarding the listener's expectations of relevance (e.g. *My watch stopped working and I need to re-calibrate it, could you tell me what time it is?*) participants consistently gave more precise answers, presumably because they considered this to be relevant information for re-calibrating a watch (which was the explicitly stated intention of their interlocutor).

*14) Sebastian's cat is a princess.*

A metaphor such as (14) is another case of simultaneous broadening and narrowing of the lexical meaning. The ad hoc category PRINCESS*, meaning roughly "spoiled and needy", is constructed by including all things that are spoiled and needy but are not literal princesses (such as Sebastian's cat) and excluding all literal princesses that are not spoiled and needy.

This leaves us with an account of metaphor comprehension that is more specific and context-sensitive than the dual reference account regarding the sub-processes involved in comprehension and how they interact, even though the output of the process remains the same (the construction of an ad hoc category).

The account is also more specific regarding the parameters that facilitate lexical modulation. Instead of these being generated by the salient features of the metaphoric topic alone, they are a product of the set of interpretative assumptions that the listener considers to be relevant. This can explain the differences in meaning construction between sentences (9a) and (9b).

## 2.3 Metaphor as Implicit Comparison

A different approach in the study of metaphors to the one introduced in the previous sections originates from thinking about how a metaphor highlights relational commonalities between concepts. Whereas categorization approaches view sentences such as (14) as primarily stating a property of Sebastian's cat, it is also possible to think about what this sentence says about princesses in general. Why would someone wish to use princesses to conceptualize a cat? What is it about princesses that allows for such a conceptualization to meaningfully take place? Another way to ask these questions is to think about the relationship between sentence (14) and (15) below.

*15) Sebastian's cat is like a princess.*

Sentence (15) is a simile that explicitly draws a comparison between the cat and the princess. Arguably, it conveys the same information as (14), namely that there is some relevant point of comparison between cats and princesses that can be highlighted. When seen this way, metaphors, similes analogies and other forms of comparison seem to fall under the same category of processes that establish correspondences between situations, concepts and sets of correspondences. This is the main idea behind theories of metaphor as **Implicit Comparison**.

The most thoroughly developed instantiation of such a theory is that of Structure-Mapping Theory (also referred to as analogical reasoning), developed by Dedre Gentner and collaborators(Gentner & Markman, 1997; Gentner et al., 2001; Gentner & Bowdle, 2008; Wolff & Gentner, 2000; Wolff & Gentner, 2011, among others). However, there are other theories that share the idea that metaphors involve mappings across domains. Two very prominent instances are Conceptual Metaphor Theory (CMT) (Lakoff, 2008; Lakoff & Johnson, 2008) and Blending Theory (Fauconnier & Turner, 2003), both developed in the tradition of cognitive linguistics.

CMT puts forth the idea that individual metaphoric expressions are mere surface manifestations of a network of pre-existent mental mappings known as Conceptual Metaphors, whose function is to allow us to understand abstract domains in terms of concrete ones. According to Lakoff and collaborators, metaphors are not a language phenomenon, but fundamentally a phenomenon of the domain of thought. Thus, when we hear metaphoric expressions such as *Pete and I have been going in different directions for years* or *Our relationship hit a dead end*, an initial step in understanding them is the activation of the conceptual metaphor LOVE IS A PATH (where LOVE is called the "target domain" and PATH the "base domain"), that these two sentences are believed to be instantiations of. A main tenet of CMT is the belief that metaphors are not special uses of language that require additional mechanisms. Instead, they are part of everyday life and can be explained by the same mechanisms as other instances of language. This is a common feature of all the contemporary theories of metaphor comprehension discussed in this chapter (and emerged as a response to Grice's account). What makes CMT particularly akin to structure-mapping is the way that it posits that metaphor understanding requires the alignment of structural relations across the elements of a metaphor.

This happens via the requirement that the target domain in a metaphor conserve the cognitive topology of the base (Lakoff, 1990).

Whereas CMT studies the ways in which metaphors can shape cognition and society, Blending Theory's main concern is to examine how meaning is dynamically constructed. In the domain of metaphors, the main focus is to understand how a new mental representation arises from the combination (or "blend") of topic and vehicle. For this purpose, they posit an intricate network of input spaces that contain the conceptual features of each element individually (*cat* and *princess*, in (15)) and a blended space, which holds a superimposed mental representation of both elements simultaneously (a "cat-princess", or "princess-cat"). Multiple mappings are then allowed to happen between the different spaces (Fauconnier & Turner, 1998, 2008). As with CMT, Blending theory is related to structure-mapping in that an underlying process of structural comparison is a requirement for systematic mappings across domains.

The way in which CMT, Blending Theory and Structure-Mapping can be reconciled has been extensively discussed. See, for example, Gentner et al. (2001); Holyoak and Stamenković (2018); and Murphy (1996) on the compatibility between CMT and structure-mapping, and Coulson and Oakley (2005) for a suggestion on how Structure-Mapping Theory can be integrated into Blending Theory. Given that Structure-Mapping has been spelled-out in terms of a cognitive architecture with specific mechanisms that are empirically testable, I will take it as the prototypical representative of Implicit Comparison views.

### 2.3.1 Structure-Mapping Theory

In Structure-Mapping Theory, concepts are assumed to be organized as information about objects and their properties, relations between objects as well as higher-order relations between relations. When processing a metaphor, these networks of interconnected information are systematically mapped between the elements of the metaphoric expression: A system of relations that holds for one metaphoric element is understood as also holding for the other by means of **reasoning analogically** about the systematicities in the structure of concepts.

Analogical reasoning is spelled out as an online processing theory (Gentner et al., 2001) and has also been computationally instantiated as the Structure Mapping Engine (SME) (Falkenhainer et al., 1989). According to Structure-Mapping Theory, a metaphor is understood in two stages: An initial stage of structural alignment and a secondary one of projection of inferences from vehicle to topic (Gentner & Bowdle, 2008). The initial stage consists of establishing an explicit set of correspondences between the relational structure of the two elements in a metaphor. This stage is itself divided into three phases in which matches between identical properties of topic and vehicle are established and then grouped together into structurally consistent clusters called "kernels". The kernels are then merged into a few structurally consistent interpretations. These interpretations are then quantified in terms of the number of partial matches between structures.

Importantly, structural alignment is shaped by specific constraints of structural consistency (Gentner et al., 2001): (1) There must be a one-to-one correspondence between the elements being mapped (these could be properties or relations that hold for both vehicle and topic); (2) parallel connectivity, meaning that the arguments of all the predicates that correspond must also correspond.

A critical feature of this model is that, throughout processing, the search for commonalities is guided by a **systematicity principle** (Gentner et al., 2001), stating that matches that are connected via "deep" higher-order constraining relations (such as relations of causality) should be preferred over "superficial" matches that are independent of one another (such as color and shape).

The second stage of processing (the projection of inferences) is also guided by the systematicity principle: Comprehenders don't project random properties of vehicle to topic. Instead, they import information from vehicle to topic that can fill missing pieces in the topic's structure according to the shared system of relations established during structural alignment (Bowdle & Gentner, 1997; Clement

& Gentner, 1991). These structural inferences are what eventually become evident in the final interpretation of the metaphoric expression.

### 2.3.2 The role of context in Structure-Mapping Theory

Structure-Mapping Theory (and its computational implementation, the SME) have three important characteristics regarding the relationship between context and a metaphoric expression that have repercussions for the way in which metaphors are processed (Gentner et al., 2001).

Firstly, the SME begins blindly. This means that it does not need to know the point of the comparison between vehicle and topic prior to computing the metaphor's meaning, which is inferred solely on the basis of the similarities in structure of both terms. This means that the relationship between context and a metaphoric expression is quite different to that posited by categorization theories (specifically by Relevance Theory), which state that context guides the derivation of the relevant features of a topic that set the parameters for lexical modulation: If the relevant features are highlighted in advance, determining the parameters for modulation will be easier, impacting the earliest stages of processing. Analogical reasoning, on the other hand, does not need a context that highlights the relevant features of the topic in order to generate a structural alignment: A context that highlights relevant features of the topic can only influence secondary processing stages, i.e. when inferences are projected.

However, this doesn't mean that context does not play any role at all. According to Structure-Mapping Theory, context can influence the early stages of processing by facilitating access to the mappings across domains that are normally established during the first phase of structural alignment: Such a context can be given by providing instances of other metaphors that make use of the same alignments of conceptual structure. This should result in a priming effect on the processing of subsequent related metaphors (Gentner & Bowdle, 2008; Keysar et al., 2000; Thibodeau & Durgin, 2008).

These multi-expression metaphors are known as **extended metaphors**. Below is an excerpt of Hector Lavoe's song "Periódico de Ayer", which provides a great example of an extended metaphor:

> *Tu amor es un periódico de ayer*
> *Que nadie más procura ya leer.*
> *Sensacional cuando salió en la madrugada.*
> *a mediodía ya noticia confirmada,*
> *Y en la tarde materia olvidada.*
>
> "Your love is yesterday's newspaper:
> no one bothers reading it anymore.
> It was sensational when it came out at dawn.
> By noon it was a well-established fact.
> In the evening it had already been forgotten."
>
> — Hector Lavoe, Periódico de ayer

The conceptualization of a relationship as yesterday's news is established at the beginning of the song and carried on across multiple lines (in fact, throughout the majority of the piece). According to Implicit Comparison accounts, the subsequent lines should be easier to process metaphorically because the conceptual domains have successfully been aligned in the first line of the song, so the initial stage of alignment in subsequent metaphors can be easily sped-up.

A second characteristic of the SME is that it can derive multiple interpretations from the same metaphor based on the different number of possible structural alignment matches between the elements of the metaphor. This has some interesting theoretical consequences. The first one is intuitively true: Sometimes a single metaphor can generate multiple meanings (see for example the discussion of sentences (9a) and (9b) earlier in the chapter).

Generally, the SME chooses between competing interpretations by counting the number of structural matches and following the systematicity principle. However, as examples (9a) and (9b) show, sometimes systematicity in the internal structure of the words by themselves is not enough to select the correct metaphoric interpretation. For this purpose, integration with both world knowledge and context are crucial. This is briefly mentioned by Gentner et al. (2001), who state the following: *"Which* *interpretation (of a metaphor) you choose will depend on (a) which has the largest and deepest common structure, as discussed above; (b) which best fits other knowledge about the target; and (c) which is more relevant to the current context"* (p. 237). But when and how does context relevance come into play? If SME aligns structure between individual concepts, it seems like context would have to come into play at later stages, since the earlier moments of processing in Structure-Mapping Theory are blind by design. This remains unclear.

Despite the uncertainty regarding the way in which context helps select the best possible interpretation, the fact that the SME generates multiple interpretations of an utterance simultaneously has an interesting theoretical consequence: It means that a sentence can have both a literal and a metaphoric interpretation, such that sentence (14) could generate the interpretation that Sebastian's cat is a member of a royal family. This is a characteristic that Structure-Mapping Theory shares with categorization views: There is no primacy of the literal meaning. Figurative and literal language are processed by the same mechanisms and in similar time-frames. This is consistent with the evidence against the Gricean system discussed earlier in this chapter (e.g. Gibbs, 1986; Glucksberg et al., 1982).

The third important characteristic of the SME that has repercussions on how metaphor processing interacts with context is the following: According to Structure-Mapping Theory, inferences happen as a natural consequence of structural alignment without the need of additional machinery. Clusters of structure are carried over from topic to vehicle in order to provide meaningful content to the metaphoric topic in a way that maximizes structural consistency. As a consequence of this, understanding a metaphor is a necessarily symmetric process in which both terms contribute equally to processing (Wolff & Gentner, 2011, p. 1459). This carries the underlying assumption that both topic and vehicle could be integrated in a similar way with the preceding context, since their relative order should not impact the way that the metaphor is understood.

I will turn to this issue in more detail in the following section, where the most prominent tests of the predictions made by theories on metaphor comprehension will be addressed, together with a discussion of their limitations.

## 2.4 Empirical Testing Grounds for Theories of Metaphor Comprehension

### 2.4.1 Role of the literal meaning

An issue of importance for metaphor theories is the role that the literal meaning of a metaphoric vehicle plays during processing:

*16) The journalist's opinion was fenced in after the change in regime.*

In (16), the verb *fenced in* entails the concept of physical containment; its direct object is something that is not allowed to physically move. However, when we hear that *the opinion has been fenced-in*, the feature of a physical barrier is not part of the final metaphoric meaning: We do not generate the interpretation that an opinion cannot physically move, but instead that the journalist is not allowed to speak freely. Physical containment is in this case a type of **literal feature** whose meaning is not carried over to the metaphoric interpretation in its original form. Similarly, in (15), several properties of the encoded meaning of *princess* are very noticeably not part of the overall metaphoric meaning: We do not believe the cat to be of noble origin, nor do we think that the cat is wearing a royal dress or a tiara.

The role of these literal features during processing is one of the battlefields of theories on metaphor comprehension, and the two main theories that I've discussed so far make very different predictions in this regard.

From a category inclusion perspective, the noun *opinion* in (16) provides the parameter of [+ abstract] (and the topic in (15), the cat, provides the parameter [+ domestic animal]). This parameter, together with the relevant utterance context, determines the way in which the metaphoric vehicle is modulated. This can be interpreted as meaning that only those features that are needed for the construction of metaphoric meaning (i.e. features compatible with the parameters provided by the topic and the discourse context) are activated, while those incompatible with the parameters are suppressed.

Support for this view comes from priming experiments. Gernsbacher et al. (2001) showed participants either a metaphoric or a literal sentence as a prime (*That defense lawyer is a shark* or *That large hammerhead is a shark*) and then asked them to perform a verification task on a sentence describing a feature of the vehicle that was irrelevant or relevant for the construction of the metaphoric meaning (*sharks are good swimmers* or *sharks are tenacious*). They found that, after reading metaphorical primes, participants were faster at verifying sentences describing a relevant feature for the metaphoric interpretation compared to when they read a literal prime. They also found that verifying sentences about a metaphor-irrelevant property took longer after reading a metaphor than after reading a literal statement. They interpreted these results in terms of activation of relevant features and suppression of irrelevant ones: When the word *shark* is used metaphorically, features such as “vicious” are enhanced and features such as “good swimmer” are inhibited. This result suggests that, after a metaphor has been understood, irrelevant literal features are no longer active, given that the meaning of the vehicle has been successfully modulated.

Rubio Fernandez (2007) conducted a similar study with the key difference that the target was a single word and it was shown at varying intervals. She found that at short intervals (0 milliseconds and 400 milliseconds) literal features that were not relevant for the metaphoric interpretation were primed by the metaphor and only actively suppressed when presented 1000 milliseconds after the prime. This suggests that these irrelevant literal features are indeed initially activated after a metaphor has been comprehended and are only suppressed at a later point.

McGlone and Manfredi (2001) deployed a reversed version of this paradigm and showed participants irrelevant (*sharks can be blue*) or relevant (*sharks can be ruthless*) literal features as primes and then metaphorical sentences (*some lawyers are* *sharks*) as targets. They found that relevant features facilitated whereas irrelevant features hindered comprehension compared to a baseline condition without a prime, suggesting that literal properties of the vehicle are suppressed early on during processing.

From the perspective of Implicit Comparison, on the other hand, the activation of different types of features of the meaning vehicle is not contingent upon parameters provided by the topic. According to Structure-Mapping Theory, the structures of the metaphoric elements have to be thoroughly scanned for similarities during the first stage of processing (Gentner et al., 2001). This forcefully requires all features of both vehicle and topic to be initially activated, because in order to find similarities in structures, it is necessary to first activate all knowledge of both structures individually.

Support for this view comes from an event-related potential (ERP) version of the metaphor paradigm. Weiland et al. (2014) showed participants a masked prime (presented on screen for only 67 ms) consisting of a word representing a feature associated only with the literal meaning of an upcoming metaphoric vehicle and irrelevant for the construction of metaphoric meaning (*furry*, for the metaphor *These lobbyists are hyenas*) followed by the metaphor itself. They found that the N400 effect (computed as the difference in stimulus-related average electrical responses between the metaphor and a sentence in which the vehicle appears in its literal meaning, such as *These predators are hyenas*) was reduced when participants saw the literal prime compared to when they did not see any prime at all. This suggests that literal features can indeed ease comprehension of a metaphor, even when they are completely irrelevant for the constructed metaphoric meaning (*furry* is not generally part of the final interpretation of the metaphor *my lawyer is a hyena*). This result is incompatible with that of McGlone and Manfredi (2001), who had found that literal features irrelevant for the metaphoric meaning (*sharks can be blue*, for the metaphor *some lawyers are sharks*) hindered comprehension of the metaphors.

The Implicit Comparison View also makes explicit predictions regarding the activation pattern of literal features after the metaphor has been understood.

They claim that, once a metaphor has been understood, the pattern of mappings between topic and vehicle, which arises as a consequence of structural alignment, remains activated and is available for subsequent processing (Gentner et al., 2001), allowing for further mapping from one domain to another. This is referred to as "incremental mapping", and has been formalized computationally as the Incremental Structure-Mapping Engine (Forbus et al., 1994). Evidence for this view comes from investigations on extended metaphors. Gentner and Boronat (1992) had participants read question-answer pairs where the answer was always identical (*No, she was doing a slow simmer*) and the question either shared the same metaphoric mapping with the answer (*Was Anna still boiling mad when you saw her?*) or had a different type of mapping (*Was Anna still a raging beast when you saw her?*). Participants read the answer significantly faster when the metaphoric mappings were shared across question and answers compared to when they weren't. It has also been found that novel metaphors prime subsequent novel metaphors that share the same conceptual mappings between domains (Keysar et al., 2000) and even that conventional metaphors can prime subsequent novel metaphors (Thibodeau & Durgin, 2008).

These findings are somewhat challenging to account for from the perspective of Category Inclusion, which seems to posit that metaphor comprehension occurs only locally: If the meaning of the metaphoric vehicle is altered so that irrelevant literal features are suppressed, how can these features be re-activated to prime subsequent related metaphors? One answer, coming from within Relevance Theory, is given by Carston (2010). She claims that, in an extended metaphor, the multiple related words that are semantically associated are mutually reinforcing, resulting in an enhanced activation of the literal meaning (which she calls the "lingering" of the literal meaning). This can lead to the entire literal meaning of the extended metaphor to be meta-represented and considered as a sort of "imaginary world", where the individual metaphors are understood literally. This leads to metaphoric meaning only being derived in later stages of processing (Rubio-Fernández et al., 2016). Consider the extended metaphor in Hector Lavoe's song, presented in the previous section. If hearing only the line *Your love is yesterday's newspaper*, a listener might locally adjust the expression *yesterday's newspaper* to create an ad hoc category meaning something along the lines of "a past event that is now irrelevant". However, when the song continues, the semantically related words *reading*, *sensational* and *fact* reinforce each other, so that the listener imagines a world in which love is literally the newspaper from yesterday. Further metaphoric meaning is derived later by reasoning about the similarities between the imaginary world and the real world.

Regarding the activation of literal features, the difference between Implicit Comparison views and the view of Carston (2010) seems to be that Carston (2010) would predict a facilitation effect of metaphors on subsequent related metaphors based on semantic reinforcement of related words, whereas Gentner et al. (2001) predicts a general activation of structural mapping patterns after any metaphor has been activated. In other words, the Implicit Comparison view predicts that literal features of a metaphor remain active after a metaphor is understood because these are part of a complex network of mappings between the encoded meanings of the metaphoric topic and the metaphoric vehicle. The Category Inclusion view of Carston (2010), on the other hand, predicts activation of the encoded semantic features of a metaphoric vehicle (i.e. "lingering" of the literal meaning), not of a network of systematic mappings.

To sum up, the empirical findings discussed in this section show a muddled view on the role that literal features play during metaphor comprehension. Whereas it has been suggested that literal features, when activated before the metaphor is read, hinder subsequent processing (McGlone & Manfredi, 2001) and are immediately suppressed after the metaphor is understood (Gernsbacher et al., 2001), others claim that these features can actually ease subsequent processing of a metaphor when activated prior to metaphor presentation (Weiland et al., 2014), active for at least 400 milliseconds after processing (Rubio Fernandez, 2007), and facilitate processing of subsequent related metaphors in the setting of an extended metaphor (Gentner & Boronat, 1992; Keysar et al., 2000; Thibodeau & Durgin, 2008). Outside of the setting of an extended metaphor however, these features should not remain activated (which can be inferred from Carston, 2010). This lack of agreement in the literature shows that there is need for further evidence in this debate, given its repercussions for theory development.

### 2.4.2 Are metaphors processed symmetrically?

A further critical aspect in which the theories differ is the role that the topic and vehicle play during the construction of metaphoric meaning. The implicit comparison view sees both elements playing the same initial role. In Structure-Mapping Theory, aligning the structures of the elements of a metaphor requires the same mechanisms to apply to each of them individually and then to the two of them combined: Structurally consistent kernels are found and then grouped together, forming large clusters of shared structure. This means that processing is fundamentally **symmetric**: The order of the elements does not matter, since the goal of processing is just to determine what the similarities are in the structure of the elements. Only in later stages of processing, when inferences are projected from vehicle to topic, does processing appear to be directional. However, this is only an epiphenomenon of the way in which structure-mapping attempts to fill missing holes in the structure of the metaphoric topic (Clement & Gentner, 1991). In practice, this means that the sequential order of the elements should not be a determining criterion for whether or not a comprehender can derive meaning from a metaphoric expression.

The Category Inclusion view, on the other hand, assigns role-specific tasks to topic and vehicle (e.g. Glucksberg, 2001, p 55-56): The topic's job is to provide parameters for modulation. It guides the interpretation of the upcoming stream of information by informing the comprehender of the relevant set of dimensions that are going to be necessary to process the upcoming metaphoric vehicle. Upon encountering the vehicle, an ad hoc category is constructed on the basis of the vehicle's encoded meaning and the dimensions provided by the topic and the contextual relevance constraints. If the vehicle were to be encountered prior to the topic, processing would necessarily have to be different: Deprived of all the necessary parameters for lexical modulation, the vehicle's encoded meaning would have to be activated in full, only to be reassessed later, after encountering the metaphoric topic.

Wolff and Gentner (2011) addressed this issue experimentally by examining the comprehension times of reversed metaphors: They showed participants highly directional nominal metaphors (*a rumor is a virus*), their reversed counterparts (*a virus is a rumor*), literal category statements (*the apple is a fruit*) or scrambled statements (*the cat is a library*). Participants were asked to quickly give a comprehension judgment (in the form of a yes/no answer) after the sentence had been presented for 1200 or 1800 milliseconds (in Experiment 1); 600 or 1200 milliseconds (in Experiment 2); and 500 or 1600 milliseconds (in Experiment 3). The reasoning was as follows: If initial stages of metaphor processing are symmetric, then forward and reversed metaphors should be comprehensible at comparable rates in the earliest time windows, whereas literal category statements should consistently be judged as more comprehensible and scrambled statements as more incomprehensible than the metaphors in every time window.

The results confirmed their hypothesis and showed that at the earliest time windows (500 and 600 milliseconds) participants gave comparable judgments to the metaphors and reversed counterparts. The literal category statements were deemed to be significantly more comprehensible than both types of metaphoric, while the scrambled statements were deemed to be the least comprehensible of all. The authors argue that this is evidence in favor of an initial alignment stage in which both topic and vehicle are evaluated equally regardless of their position: If during initial stages of processing the order of the metaphorical elements mattered, then we should have seen significantly lower comprehensibility ratings for the scrambled compared to the regular metaphors.

However, this evidence is only indirect, since it is based on the absence of an effect: The fact that participants were similarly likely to rate forward and reversed metaphors as comprehensible is not an unequivocal indicator that they are being processed equally. This is an important point, seeing that forward and reversed metaphors can have diametrically different meanings, as evidenced by sentences (17) and (18).

*17) My butcher is a surgeon*

*18) My surgeon is a butcher*

This stark difference in meaning makes it difficult to draw direct conclusions on comparable comprehension rates: It is simply not felicitous to compare these types of forward and reversed metaphors. Wolff and Gentner (2011) also put participants through an extensive practice round prior to testing: Participants went through 64 practice trials (which included metaphors and reversed metaphors) prior to the actual 72 experimental items. This makes it possible that participants developed a task-specific comprehension strategy. That being the case, the results could be seen as a reflection of this strategy and not of how metaphors are normally comprehended.

A further important point pertains to the role of context. Wolff and Gentner (2011) tested the hypothesis of symmetry on nominal metaphors presented individually. However, as examined previously in this chapter, categorization views (Relevance theory in particular) highlight the role of context for the construction of metaphoric meaning: The parameters for modulation are not provided by the topic exclusively but by a combination of topic, a listener's assumptions and background information, as well as the relevant discourse context. A more ecologically valid test of the theories would have been to provide participants with a brief context before every forward and reversed metaphor.

### 2.4.3 Mediating factors: conventionality, aptness and familiarity

The final issue to be discussed is the possibility that metaphor comprehension is not a unitary process and that different processing routes could be available depending on underlying mediating factors such as grammatical form, individual differences between comprehenders, semantic properties of the stimuli or the influence of context.

In fact, the literature on metaphor processing has provided evidence for various factors having a mediating effect on comprehension. Three haven been given particular importance throughout the years: the first one, **metaphoric aptness**, is defined as the degree to which the figurative meaning of a metaphoric vehicle captures important features of the metaphoric topic (Jones & Estes, 2006). Sentence (19) could be said to be fairly apt: The property "ruthless" of the lexical item *shark* is made salient (since this is one of the only few properties of sharks which can be used to describe the personality traits of a human being, as opposed to "having fins" or "living underwater") and it can capture (arguably) relevant features of the topic *my landlord* with relative ease (since in some contexts the word *landlord* has a negative connotation and it is associated with ruthless individuals, see for example the German word *Miethai*). The second key factor that mediates comprehension is **metaphor conventionality**, which is determined by the frequency in which a metaphoric vehicle is used in its metaphoric meaning in a given language (Bowdle & Gentner, 2005): The metaphoric vehicle in sentence (19) (*shark*) might be frequently used in this particular way, so (19) is a fairly conventional metaphor. The third mediating factor is conceptually related to conventionality and tries to capture the frequency of use of the entire metaphoric expression, i.e. topic and vehicle, and is known as **metaphoric familiarity** (Thibodeau & Durgin, 2011): Whereas sentence (19) might be highly familiar, (20) below might not be, even though their conventionality would be thought of as being the same:

*19) My landlord is a shark.*

*20) Eva is defending a shark. [in a context in which she is a landlord's lawyer]*

Both of the leading views on metaphor comprehension – Implicit Comparison and Category Inclusion – state that different metaphors might follow different processing routes during comprehension. The key difference between the theories in this regard is which one of the three factors (aptness, conventionality or familiarity) is viewed as responsible for the shift between processing routes.

Category Inclusion views state that in most cases metaphors will be understood via category inclusion. However, there are instances in which such a strategy will fail and language users will resort to analogical reasoning in order to derive meaning from a metaphoric expression. Specifically (in what he informally referred to as the quality-of-metaphors hypothesis), Glucksberg (2008) states that metaphors with low aptness values might not be easily understood via category inclusion and will instead most likely be understood as an implicit comparison through analogical reasoning. This is a consequence of the way in which categorization is thought to operate: If the topic and discourse context provide parameters that are not compatible with the retrieved features of the metaphoric vehicle it will require a deeper analysis of the structure of both elements, thus engaging in the first stage of analogical reasoning: structural alignment.

Implicit Comparison views, on the other hand, see analogical reasoning as the default mechanism applied to any given metaphoric expression. However, this can be modulated either by a metaphor's conventionality (Bowdle & Gentner, 2005) or familiarity (Thibodeau & Durgin, 2011), which both refer to frequency of exposure to either the vehicle (conventionality) or the topic-vehicle pair (familiarity). The reasoning is as follows: Through repeated exposure, the literal meaning of a metaphoric vehicle will become progressively more associated to its figurative meaning and will more likely be processed as a category statement. Eventually, the metaphoric vehicle will become completely fossilized and its meaning will be merely retrieved from memory (such as the word *legs* when speaking of a chair or a table). This view is known as the "career of metaphor hypothesis" (Bowdle & Gentner, 2005): When hearing the word *shark* used metaphorically to describe a landlord for the first time, it will necessarily require analogical reasoning. However, the more often this specific metaphoric vehicle (according to Bowdle & Gentner, 2005) or this specific topic-vehicle pair (according to Thibodeau & Durgin, 2011) is heard, the more likely it is that it will be processed as a category statement given that repeated exposure strengthens the association between the vehicle's literal and figurative meaning. Eventually, the word *shark* acquires the secondary meaning "ruthless" and it is processed in the same way as a word with multiple stored meanings such as *bank* or *bug*.

The influence of familiarity, conventionality and aptness on metaphor processing has been shown experimentally on numerous occasions (Blasko & Connine, 1993; Bowdle & Gentner, 2005; D. Chiappe & Kennedy, 1999; D. Chiappe, Kennedy, & Chiappe, 2003; D. Chiappe, Kennedy, & Smykowski, 2003; Jones & Estes, 2006): More apt metaphors are understood faster and better than less apt ones, more conventional metaphors are understood faster and better than less conventional ones (Gentner & Wolff, 1997), and more familiar metaphors are understood faster than less familiar ones (Thibodeau & Durgin, 2011). What does this tell us about the comprehension mechanisms involved?

Bowdle and Gentner (2005) investigated this by examining the differences in comprehending metaphors and similes. Their logic was as follows: Similes (*my lawyer is like a shark*) explicate a comparison between two terms and trigger analogical reasoning, whereas metaphors have the surface form of Category Inclusion and are more likely to be understood as category statements. Thus, by examining which type of metaphors people prefer in the simile or category statement form, it is possible to draw conclusions about the processing route taken: analogical reasoning or category inclusion. They showed that (1) less conventional metaphors were preferred in the simile form compared to more conventional ones, which were preferred in the category statement form (Experiment 1); (2) less conventional metaphors were read faster in the simile vs category statement form, whereas more conventional metaphors showed the opposite pattern; and (3), when repeatedly exposed to the figurative meaning of a metaphoric vehicle throughout an experiment, participants then showed a preference for new metaphors using this same vehicle to be formulated as category statements and not as similes. Because of this, the authors concluded that conventionality of a metaphoric vehicle determines whether a given metaphoric expression is processed through analogical reasoning or through Category Inclusion.

A problem with deciding whether conventionality or aptness is the true mediator of processing route is the fact that these two factors are often correlated. In fact, they were correlated in the materials used by Bowdle and Gentner (2005). This seems to be the case in several other studies on metaphor processing as well: Jones and Estes (2006) found high correlation values between aptness and conventionality for 18 different published experiments, leading them to question the interpretation of Bowdle and Gentner's results. To address this, they created a new set of materials in which conventionality and aptness where completely orthogonal and used these materials to re-run Bowdle and Gentner's (2005) experiments. They found that, in all three experiments, aptness and not conventionality accounted for the form preference and reading time differences reported by Bowdle and Gentner. Their conclusion was that, in line with categorization views, how good a metaphor is, and not its conventionality, predicts whether people will process it as an implicit comparison or as a category inclusion statement.

A counter-argument to Jones and Estes' (2006) conclusion was later provided by Thibodeau and Durgin (2011). They argued that familiarity is a third mediating factors needs to be considered in this discussion. According to them, familiarity and not aptness is the driving factor that will determine the mechanism through which a metaphor will be understood. This could be interpreted as an addendum to the Career of Metaphor Hypothesis: Through repeated use of specific topic-vehicle pairs, a vehicle's literal meaning strengthens conceptual links to its metaphoric meaning, eventually leading to the fossilization of the figurative meaning of the vehicle.

## 2.5 Going Forward

The points raised in this chapter should make it clear to the reader that the debate on metaphor processing is far from resolved. Specifically, the three key issues raised - the role of the literal meaning, the symmetry of processing and the effect of mediating factors - deserve a closer examination. One way to move the debate forward would be to consider what the opposing theories do have in common. Both Implicit Comparison Views and Category Inclusion Views across the board state that metaphor comprehension does not require any additional machinery than the one routinely used to comprehend literal language (Coulson & Oakley, 2005; Gentner et al., 2001; Glucksberg, 2001; Lakoff, 2008; Sperber & Wilson, 2008). So if metaphors aren't special, why don't we approach the study of metaphor comprehension the same way that the study of literal language has been approached?

In this regard, two points are worth mentioning:

1) Regardless of theoretical inclination, a common characteristic of metaphor research has been to examine the processing of metaphors both in isolation (i.e. without any context) and only once the entire metaphoric expression has been presented and understood. This is particularly noticeable in the work that has been the most important for theory development (e.g., Bowdle & Gentner, 2005; Gernsbacher et al., 2001; Glucksberg et al., 1997; McGlone & Manfredi, 2001; Rubio Fernandez, 2007; Wolff & Gentner, 2011). However, research on language-vision interactions and on incremental language processing (see for example Knoeferle, 2015; Tanenhaus et al., 1996) has shown that understanding language involves an ongoing and simultaneous integration of different types of information (visual and linguistic, among others) in order to update our beliefs and our mental representation of a given event, as well as to generate expectations about incoming (linguistic) input. Expectations about upcoming linguistic information can be guided by (1) possible referents that have been established in the previous linguistic context (Altmann, 1999), (2) the visual context (Knoeferle et al., 2005), (3) semantic and world-knowledge constraints imposed by the head of a phrase, such as verbs (Ferretti et al., 2001), and the strengthening of these constraints via available visual referents (Altmann & Kamide, 1999), among others. Thus, in

order to study the contribution of the individual elements of a metaphor to the overall interpretation during real-time processing, it would be immensely beneficial to consider the way in which processing unfolds over time as well as how it interacts with the context in which the metaphoric utterance is embedded.

2) Metaphor research aimed at theory development has focused almost exclusively on English nominal metaphors. Holyoak and Stamenković (2018) conducted a large-scale review of studies on metaphor comprehension. They surveyed a pool of 4900 articles on metaphors published in scientific articles since 1976. Of these, they selected 77 that exclusively dealt with testing theoretical accounts of metaphor comprehension on healthy adult participants. The conclusions of the authors are damning:

> *Perhaps the most obvious limitation of much of the work on metaphor over this review period is its unbalanced focus on nominal metaphors when testing between theories. (...) Besides advocating investigation of a broader range of metaphors, we would also call attention to the need to broaden the range of languages and cultures in which metaphors are investigated.* (Holyoak & Stamenković, 2018, p. 643)

In a field populated by contradicting results advocating for opposing theories, it is imperative to seek out different sources of evidence. Specifically, evidence from non-nominal metaphors in languages other than English.

This dissertation tackles the three key issues discussed in this chapter by directly improving on points (1) and (2). Ultimately, the goal is to seek out empirical evidence that can help solve the long-standing debate on metaphor comprehension. The next three chapters embody this attempt: They present a series of experiments that meet the challenge raised by Holyoak and Stamenković (2018) by investigating two different types of non-nominal metaphors: Verbal metaphors and verb-object metaphors. Importantly, this investigation is carried out in German, an underrepresented language in metaphor research.

Before moving on to the empirical part of this dissertation, the last section of this chapter introduces some of the findings in language-vision interactions and incremental language processing that will set the stage for the next chapters.

## 2.6 Language-Vision Interactions and Incremental Language Processing

Sentence processing does not occur in isolation. Instead, language comprehenders actively seek out different sources of information that they can integrate together in order to update their previous beliefs and construct a more complete picture of an event, a situation, or a speaker's communicative intention. This idea, fundamental to Relevance Theory as described earlier in this chapter, is also essential to research on language-vision interactions (see Huettig et al., 2011; Knoeferle & Guerra, 2016; Knoeferle, 2019; Tanenhaus & Trueswell, 2006, for reviews).

The starting point for this field was the pioneering work of Cooper (1974), who showed that there is a close temporal adjacency between language understanding and the processing of visual stimuli. In the study, participants heard stories while simultaneously being presented with images of potential referents while their eye movements were monitored, something that years later came to be known as the Visual World Paradigm (VWP) (see Huettig et al., 2011, for a comprehensive review of the method). The results of this study showed that participants looked at the visual representations of objects immediately after they were mentioned in a story, highlighting the rapid way in which language and visual processes interact.

It wasn't until the mid-1990s that the VWP started to be widely used by psycholinguistic researchers. A team of researchers from the University of Rochester lead by Michael Tanenhaus began studying the interaction of language and visual perception to answer questions about syntactic ambiguity resolution (Tanenhaus et al., 1995), the role of phonology in lexical access (Allopenna et al., 1998), the influence of prosody on sentence comprehension (Dahan et al., 2002), among others.

The VWP has successfully been deployed to answer question about the type of expectations that develop during language comprehension and how these can guide the processing of incoming linguistic stimuli. This line of research, developed primarily by Gerry Altmann and collaborators, focuses on using the visual context as a reflection of what participants in an experiment anticipate that a linguistic discourse will refer to next.

In a now classic study, Altmann and Kamide (1999) showed participants a visual scene depicting an agent and four possible objects that the agent could interact with. The authors found that when participants heard sentences (such as *the boy eats cake*) that included verbs with particular selectional restrictions (*eats* requires an object with the feature [+ edible]) participants' eye movements, upon hearing the verb, anticipated the direct object by focusing on the only object in the visual context that was compatible with the verb's selectional restrictions (a picture of a cake, as opposed to a picture of a ball, a toy car or a train set).

Kamide et al. (2003) followed-up on this finding and showed that even when a verb does not on its own provide enough evidence for participants to anticipate a verbal object, the combination of a verb and an agent can do just that: Upon hearing *the man will ride...* participants looked more at a visually represented motorbike than when they heard *the girl will ride....*

These findings support a view of language processing according to which language comprehenders construct a dynamic mental representation of an event that is updated as soon as more information becomes available, and this information is in turn used to generate expectations about the nature of further input. This is crucial for our purposes since it suggests that it is necessary to consider the incremental processing of a metaphor in order to assess the relative contribution of each one of the metaphoric elements to a comprehender's dynamic mental representation. This will be particularly important when we turn to the question of metaphoric symmetry in chapter four.

Through eye-tracking technology it has also been shown that the processing of visual stimuli interacts with the processing of written abstract language. Guerra and Knoeferle (2014) showed participants a video of two playing cards that either moved closer together or further apart. Participants then read German sentences that dealt with semantic dissimilarity such as *Frieden und Krieg sind bestimmt verschieden* ("Peace and war are certainly different") or similarity, such as *Kampf und Krieg sind freilich entsprechend* ("Battle and war are certainly similar"). Their results showed that when the motion of the cards was conceptually aligned with the direction of the semantic relation (close~similar; far~different), participants were faster at reading the second of the presented nouns (Experiment 3) as well as the adjective (Experiments 1 and 2) than when there was no such conceptual alignment. The result was interpreted as evidence for an abstract co-indexing link between spatial distance and semantic similarity. One characteristic of the eye-tracking during reading method is that it allows for a rough mapping of the results onto different stages of language processing (Clifton et al., 2007; Rayner, 1998; Vasishth et al., 2013, for a counterargument). The fact that Guerra and Knoeferle (2014) found effects in first-pass reading times (considered a measure of early stages of processing) can be interpreted as a sign of the early and rapid integration of language processing and the visual context.

It's important to note that Guerra and Knoeferle (2014) investigated the effects of the visual context on the processing of concepts that have been retrieved from memory, such as the meaning of the words "war" and "peace". But how does the visual world interact with processing concepts that are not retrieved from one's mental lexicon, but are instead constructed on the fly, such as metaphors?

This question has been previously approached from the perspective of language production. Sato et al. (2015) investigated whether showing participants images depicting spatial containment would encourage them to produce expressions in which spatial containment is used metaphorically to speak of abstract difficulty. They found that even when the sentences they produced were thematically unrelated to the images viewed, participants still produced more metaphors drawing from the domain of spatial containment than when they saw a neutral picture as prime. The authors, who work within the CMT framework (Lakoff & Johnson, 2008), interpreted the result as evidence for an activation of the Conceptual Metaphor DIFFICULTY IS CONTAINMENT after having seen the pictures, leading to the production of individual linguistic metaphors derived from this specific Conceptual Metaphor.

So if visual depictions of spatial containment can facilitate the production of metaphors that have spatial containment as part of their literal meaning, can depictions of physical containment facilitate the comprehension of containment metaphors? It is generally assumed that there is at least some overlap between mental representations accessed during language production and the ones accessed during language comprehension (Gambi & Pickering, 2017). Some theories even specify that the representations are completely identical (Pickering & Garrod, 2004) or at least coincide in their semantics (Levelt et al., 1999).

Given this background, it is worth exploring whether activating the feature of spatial containment could facilitate comprehension of metaphors of difficulty that have spatial containment as part of their encoded meanings in a way that is analogous to what has been posited by Sato et al. (2015) for language production. This would give us insight into the role of literal features during online processing of metaphoric expressions, one of the main issues discussed in this dissertation. With this in mind, I now turn to the first set of empirical studies, which examines how visual depictions of literal features affect processing of subsequent verbal metaphors.

# 3
# The Role of Features of the Literal Meaning

The studies in this chapter were designed to investigate the first of the three major issues of this dissertation, namely the role of the literal meaning of the metaphoric vehicle during processing.[1] Specifically, this chapter examines the role of spatial containment during processing of metaphors in which verbs of spatial containment are used to convey abstract difficulty. The following questions are addressed: Does activating semantic features associated only with the encoded literal meaning of a metaphoric vehicle (i.e. spatial containment) influence processing of novel verbal metaphors? And do said features remain active after a metaphor has been understood?

The theoretical views presented in chapter two provide different answers to these questions. The Implicit Comparison View states that the encoded meaning of both elements in a metaphor must be processed fully in order for their structures to be aligned (given that structural alignment is a process that begins blindly) (Gentner et al., 2001). This means that all features associated with the literal meaning of

[1]This chapter is a slightly modified version of Rodriguez Ronderos, Guerra & Knoeferle (under review). Camilo Rodríguez Ronderos, Ernesto Guerra and Pia Knoeferle conceptualized and designed the experiments. Camilo Rodríguez Ronderos and Ernesto Guerra analyzed the data. Camilo Rodríguez Ronderos wrote the article with feedback from Ernesto Guerra and Pia Knoeferle.

a metaphoric vehicle must be activated during the first stages of processing, so activating literal features prior to comprehension should facilitate processing of the metaphoric vehicle (Weiland et al., 2014). It is also likely that these features remain active after the metaphor has been understood, because the pattern of structural mappings between topic and vehicle can be used for subsequent processing, as has been shown to be the case for extended metaphors (Gentner & Boronat, 1992; Gentner et al., 2001; Thibodeau & Durgin, 2008).

The Category Inclusion View, on the other hand, states that the meaning of the metaphoric vehicle is modulated early on in processing given the parameters provided by the metaphoric topic (Glucksberg, 2008). This means that features related to the literal meaning of a vehicle that are not compatible with these parameters will be suppressed, so activating them prior to comprehension should hinder processing of the metaphoric vehicle, resulting in an interference effect (McGlone & Manfredi, 2001). These features should also remain suppressed immediately after processing has been completed (Gernsbacher et al., 2001) or at least after 1000 milliseconds (Rubio Fernandez, 2007). In an extended metaphor however, literal features can “linger” (Carston, 2010) and facilitate access to subsequent related metaphors (Rubio-Fernández et al., 2016). Importantly, this facilitation is caused by a simple activation of literal features and not by an activation of a network of systematic relations between topic and vehicle.

I conducted 4 experiments to investigate the predictions made by the theories. In Experiments 1 and 2 (eye-tracking during reading), participants saw short animated clips depicting physical containment. They then read sentences in which verbs of physical containment were metaphorically used to signify difficulty, such as in sentence (1) below, and then answered questions about either the sentences or the videos.

1. *It was difficult for the journalist to see his opinion fenced in after the change in regime.*

The two animated clips both showed a ball with identical movement. The only difference between videos is that in one of them, a box hovers over the ball and appear to trap it. The ball then struggles to escape and comes to a halt. This is meant to elicit a mental representation of physical containment, given that the most important feature of this scene is the entrapment of the ball. In the second video, the ball simply appears to move freely in the absence of a box and comes to a stop on its own, with no box to be seen. Stills from both videos can be seen in Figure 3.1).

The goal of these two experiments was to study how seeing a video depicting physical containment - which I take to be a prominent semantic feature of the encoded meaning of the verbs used in all our sentences, yet incompatible with the meaning of the individual metaphors - interacts with the processing of verbs of spatial containment used metaphorically. I compared this to how the same sentences are processed after seeing a video clip that does not share the conceptual feature of containment with the verbs. In these two experiments participants also answered questions about what they saw in the video after reading the sentence. This should provide insight on the role that literal features might play after a metaphor has been understood.

In Experiment 3 (self-paced reading), I examined how participants would naturally answer the same questions asked in Experiments 1 and 2 (after sentence comprehension) when the video clips are followed by literal sentences instead of metaphors. Doing this provided a baseline measure to interpret the results of the question-answering times of Experiments 1 and 2.

Finally, Experiment 4 (lexical-decision task) investigated how the video clips of Experiments 1-3 interact with the verbs of spatial containment from Experiments 1 and 2 when said verbs are read in the absence of a context (i.e., when participants are expected to retrieve the encoded, literal meaning only).

## 3.1 Experiment 1

I began the investigation by asking the following question: Will watching video clips of spatial containment facilitate or hinder comprehension of metaphors made up by verbs of spatial containment? Additionally, how will the activation of spatial containment interact with processing the metaphorically used verbs after the metaphors have been understood? Experiment 1, an eye-tracking during reading study, was designed to answer these questions.

### 3.1.1 Participants

Forty-eight monolingual university students who were native speakers of German (ages 18–31) were recruited and tested at the Humboldt-Universität zu Berlin. All participants were right handed and had normal or corrected-to-normal vision. They received 8 euros as compensation upon giving their written informed consent. This study was covered by the ethics vote of the psycholinguistics lab of the Humboldt-Universität zu Berlin.

### 3.1.2 Materials and design

I created 40 critical items consisting of German metaphorical sentences. All sentences had an identical syntactic structure, namely a main clause with an infinitive subject clause, as exemplified in (2). In the infinitive subject clause, a verb of physical containment, which always appeared in the same position, was used metaphorically to denote abstract difficulty. In the main clause, it was asserted that the situation described in the infinitive clause was "difficult". All critical sentences can be found in Appendix A.

(2) *Es war für den Redakteur /schwierig* ADJ */, seine / Meinung* TARGET NOUN */ nach dem Regimewechsel / umgittert* VERB */ zu sehen.*

*It was for the journalist/ difficult* ADJ */, his / Opinion* TARGET NOUN */ after the change in regime/ fenced-in* VERB */ to see*

"It was difficult for the journalist to see his opinion be fenced-in after the change in regime."

**Sentence norming**

Prior to running Experiment 1, I conducted a norming study of the target sentences to make sure our metaphors were readily understandable. A sample of 15 participants, who did not participate in the main study, were asked to rate 80 sentences on a scale of 1 to 7, with 1 being totally incomprehensible and 7 being totally comprehensible. The 80 sentences were made up of the critical 40 metaphoric sentences and 40 semantically incoherent filler sentences (e.g. *It was sad that Thomas drank the car so fast*). Order of presentation of the sentences was randomized. The goal of the norming task was to establish whether any of the critical metaphorical sentences would be rated as incomprehensible (meaning a rating of 3.5 or lower) and whether the metaphorical sentences were rated significantly higher than the semantically incoherent sentences.

**Results of the norming task**

Four of the forty critical sentences were rated lower than 3.5 on average and were dropped from the investigation. The remaining 36 sentences were used in all following experiments.

To determine whether these 36 sentences were in fact understood, an ordered logistic regression model was fitted to the data (Gelman & Hill, 2006). The model was constructed to see whether our critical items and the semantically incoherent fillers could predict the 1-7 ratings. The results show that a change from level 0 (semantically incoherent) to level 1 (critical item) was associated with an increase of odds ratio of 7.96 ($t= 17.5$, $p<0.001$) This means that for metaphorical sentences, the odds of being rated higher were 7.96 times those of incoherent sentences, holding constant all other variables. The data therefore strongly suggests that participants were able to determine a difference in meaning between the semantically incoherent sentences and the novel metaphoric sentences.

Furthermore, the ratings collected on the norming task were used to refit the statistical models of Experiments 1 and 2. These results are reported in chapter five.

#### Filler sentences

Seventy-two filler sentences were constructed to reduce the likelihood of strategic behavior and to mask the purpose of the investigation. I thus had 24 German idioms as fillers with similar syntactic structure to our critical items, as well as 24 novel metaphors different to the critical items. The remaining 24 filler sentences were literal statements.

#### Visual primes

Two critical videos were created using Adobe Photoshop. Each video showed a ball bouncing with identical motion: In one of them (used in the “contained” conditions) the ball was seen to be captured by a moving box, forcing the ball to a still stand. In the other (used in the “not-contained” conditions), the ball bounces freely and stops on its own. Figure 3.1) shows a series of stills for each of the videos.

Furthermore, inspired by Experiment 1 of Guerra and Knoeferle (2014), two versions of each video were created: One with a printed word from each critical sentence on the ball and one without any printed word. Participants thus saw, for example, a video of a box trapping a ball (or a ball bouncing freely) that had the word *opinion* written on it, and subsequently read sentence (2), in which an “opinion” is said to be *fenced in*. This was done to maximize the possibility that participants would establish a relation between the visual context and the written sentence. For the filler trials, 4 other animated videos were created that were randomly paired with the 72 filler sentences.

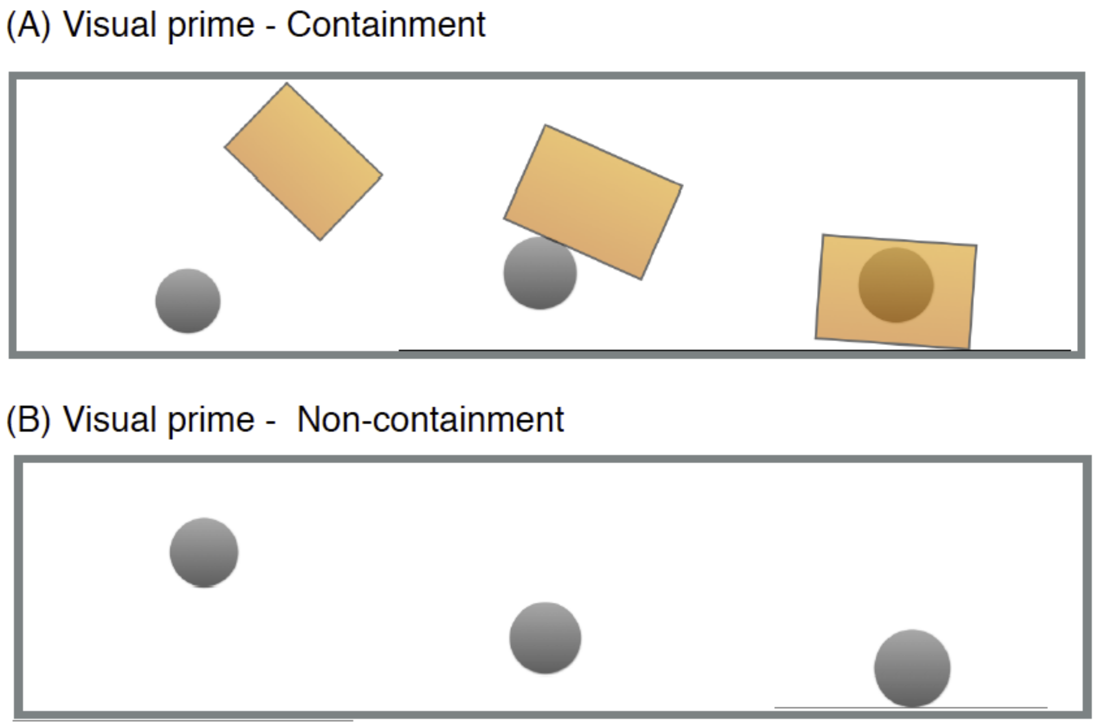


**Figure 3.1:** Stills from the video used in the 'contained' and 'not-contained' conditions of experiments 1-4

### Comprehension questions

To investigate the role of literal features after a metaphor has been comprehended, I included a comprehension question after every trial. For critical trials, the question was always about the video, either (a) referring to the ball (*Was the ball in the box?*) or (b) to the metaphoric topic that may or may not have appeared written on the ball in the video (*Was the opinion in the box?*). Trials with incorrect answers were discarded from the analysis.

The idea of having these two different questions was that they might allow us to investigate different ways in which literal features could be activated after metaphor comprehension: It could be the case that literal features are simply activated because they are seen in the video and mentioned in the sentence, in which case question (a) should be easier to respond to when the video-prime seen prior to the metaphor activates the literal feature of containment. This would be compatible with the Implicit Comparison View and with Carston's (2010) "lingering" of the literal meaning view. Alternatively, literal features could remain activated because they are part of a network of systematic associations between topic and vehicle established during structural alignment, as suggested by Gentner and Boronat (1992); Gentner et al. (2001) and Thibodeau and Durgin (2008). This would result in a facilitation effect when answering question (b), considering that question (b) suggests the parallel in structure between video and sentence by effectively "blending" together both representations. Finally, it could be the case that literal features are always suppressed, in which case neither type of question should be easier to answer when the video activates the literal feature compared to when the video does not activate it. This would be compatible with the Category Inclusion view (Glucksberg, 2008). I return to these positions and how they relate to the experimental design when discussing the predictions for the question-response times.

**Design**

Experiment 1 had a 2X2X2 Latin square design with three factors: "containment" (contained vs. not-contained), "question type" (video-question vs. noun-question) and "prime type" (no-label vs. noun-label). "Containment" refers to whether the video showed the ball bouncing freely (not-contained conditions) or being trapped by a box (contained conditions) (see Figure 1). This factor addresses the question of whether literal features can facilitate comprehension of a metaphor when they are activated before the metaphor is understood.

"Question type" refers to whether the comprehension question inquired about the video (video-question conditions) or about the metaphoric topic (*the opinion* in (2)) (noun-question conditions). This factor addresses the question of the different ways in which the literal features could remain active after comprehension of the metaphor.

Finally, "prime type" refers to whether the metaphoric topic was written on the ball (noun-label conditions) in the video prime or whether the video prime had no written language in it (no-label conditions). This factor was inspired by Guerra and Knoeferle (2014), who in their investigation of the association between physical distance and semantic similarity had the target words that appeared in the sentence also appear in the videos that participants saw before reading the sentence (Experiment 1 of their investigation). I added it as a factor to the present Experiment in order to compare whether or not there are differences in the association between literal features and their metaphoric meaning as a function of the way in which they are primed: Either by "boosting" their association to the metaphor (noun-label conditions) or by having the association be implicit (no-label conditions). Furthermore, this factor served as a type of "sanity check" for the experiment: Since the word written on the video was then read in the sentence, there should be a clear identity priming effect brought about from seeing the target noun (*opinion*) in the video and then reading it in the sentence. If we don't find such a facilitation effect, it would strongly suggest that participants were not integrating the information conveyed by the video when reading the sentence.

The dependent measures used in the experiment were three eye-tracking measures commonly associated with different temporal processing stages (Rayner, 1998) in each of the three regions of interest (i.e., the adjective, the target noun, and the verb region): First-pass reading times, defined as the duration of all fixations made in a region until the first time the region is abandoned either to a subsequent or to a prior word; Regression path duration, defined as the duration of all fixations from the first fixation in a region up to but excluding the first fixation to the right of this region; and total reading times, defined as the sum of the duration of all fixations in a region. These three measures were chosen since they can provide insight about the point in time in which effects might arise: if effects are found in first-pass or regression path duration, it would suggest that they occur during early stages of processing, whereas if they are found only in total reading times, it would suggest taht such an effect happens only during later stages of processing.

### 3.1.3 Predictions by region

The first set of predictions concerns the effect of the video on reading comprehension. I focused on analyzing three specific word regions where it would be possible for effects to arise: the adjective, target noun, and verb regions. The motivation for choosing these regions is presented in the following subsections.

### Adjective region

The motivation for expecting effects to appear in the adjective region does not come from the debate on literal features of a metaphoric expression, but is instead based on previous literature on the interaction between the visual world and written language processing. In Guerra and Knoeferle (2014) the authors found that visually depicted spatial distance facilitated reading comprehension of adjectives denoting abstract similarity. They reasoned that this facilitation effect might be due to an existing co-indexing link between spatial distance (close, far) and semantic distance (similar, dissimilar). They borrowed this idea from Conceptual Metaphor Theory, which hypothesizes the existence of such a link (Lakoff & Johnson, 2008). Conceptual Metaphor Theory also posits the existence of a link between the concepts of difficulty and containment, present in the Conceptual Metaphor DIFFICULTY IS CONTAINMENT.

Thus, to the extent in which this hypothesized Conceptual Metaphor can be accessed during online comprehension in order to process an adjective denoting difficulty (akin to the way in which a video showing spatial distance facilitated comprehension of an adjective denoting semantic similarity in Guerra and Knoeferle (2014)), watching videos of spatial containment might ease processing of said adjective. If this is the case, it should be possible to find a main effect of containment in the adjective region, with shorter reading times in the contained vs. not-contained conditions.

### Target noun region

By adding the word in the target noun region to the video (noun-label conditions), I expected a clear repetition priming effect to appear when participants encountered this word in the sentence. Concretely, if participants were able to integrate the written word from the video with the subsequently read sentence, a main effect of prime type in all dependent measures should appear, with the noun-label conditions being overall faster to read than the no-label conditions.

**Verb region**

The predictions for this region are directly derived from the debate on metaphor processing presented in chapter two. I expected a facilitation effect on an early measure, such as first-pass reading times, provided that the video relates to the literal meaning of the verb. This finding would suggest that features related to the literal meaning of a verb (in this case, physical containment) are initially active even though they might be absent from the intended metaphoric meaning. This would be in line with the results of Weiland et al. (2014), who observed that masked primes consisting of features of the metaphoric vehicle reduced the N400 effect found upon encountering the metaphoric vehicle, and would also generally support the Implicit Comparison view of metaphor understanding.

Alternatively, if activating the spatial representation of containment interferes with processing the metaphorically used verb, this should translate to longer reading times in the contained vs. not-contained conditions. This would be in line with the findings of (McGlone & Manfredi, 2001) and generally with Category Inclusion accounts that claim that literal features irrelevant for understanding the metaphor are actively suppressed during processing. Because of this, activating them should cause an interference effect. Another option is that the features are suppressed, but not to the degree in which pre-activating them via the video should cause interference. If this is the case, this could result in a null-effect of the video prime on reading the verb.

### 3.1.4 Post-sentence comprehension question

The main prediction for the response patterns to the post-comprehension questions was that if the feature of physical containment is active after participants have understood the sentence, it should be possible to find a main effect of containment on question-answering times, with overall shorter answering times in the contained vs. not-contained conditions. This would suggest that the feature of containment activated in the contained conditions (the ball is trapped by the box) was not suppressed after the metaphor was understood and facilitates answering both question (a) *Was the ball in the box?* and (b) *Was the opinion in the box?*. This would be consistent with the Implicit Comparison view.

If, on the other hand, the features activated by the video are suppressed after the metaphor has been understood, there should be either an interference or a null-effect of containment on answering question times, consistent with the Category Inclusion view.

However, because there were two types of questions, it would be possible to observe different patterns of activation of the literal feature that would bring about a more nuanced view of the role of literal features after a metaphor has been comprehended. These patterns could be important to further inform theories of metaphor comprehension. Because of this, I present here a tentative way of linking possible question-answering patterns to the theoretical views on metaphor processing, even though these links were not explicated in this way before conducting Experiment 1.

Of particular importance for a nuanced view on the role of literal features are the response times in the video-question ( (a) *Was the ball in the box?*) compared to those in the noun-question ((b) *Was the opinion in the box?*) conditions, especially in the no-label conditions (i.e. when there was nothing written on the ball seen in the video-prime): In the video-question no-label conditions, participants saw a video without a word printed on the ball and then answered question (a) above after reading the metaphor. In the noun-question no-label conditions participants answered question (b) instead.

Implicit Comparison views suggest that after a metaphor is understood, literal features remain active because they are part of the network of established mappings between topic (in this case, the target noun *opinion* in sentence (2)) and vehicle (the verb *fenced in* in (2)), which can be used to reason analogically about subsequent linguistic input. If this holds, there should be both a facilitation effect of contained vs. not-contained levels in the video-question conditions (signifying a sustained activation of the feature "containment") and a facilitation effect of contained vs. not-contained levels in the noun-question conditions (signifying a sustained activation of established mappings between different conceptual domains). This is because question (a) is a reference to the video alone, requiring only information about the feature of containment in order to answer it. If the feature is active, this should result in a facilitation effect compared to when it is not active (i.e. the not-contained condition). Question (b), on the other hand, is a complex combination of information about the sentence (given the presence of the target noun) and the video (given the reference to the box, which could have only been seen in the video), so a facilitation effect for answering this question in the contained vs. not-contained conditions would suggest that not only the feature of containment has been activated (as would be the case in the video-question conditions), but also its relationship with the metaphoric topic (the target noun).

Carston (2010), on the other hand, suggests that literal features might "linger" after a metaphor has been understood, but only as semantic features, not as part of a network of systematic associations between topic and vehicle. In extended metaphors, literal features will reinforce each other and trigger a meta-representation of the literal meaning, but not otherwise. That being the case, there could a facilitation effect of contained vs. not-contained video-prime on the question-response times in the video-prime no-label condition, but this should not be the case for the response times in the noun-prime no-label conditions.

### 3.1.5 Procedure

Participants' eye movements were recorded using an Eyelink 1000 plus desktop head-stabilized tracker, produced by SR Research. At the beginning of each experimental session, the eye-tracker was calibrated with a 9-point calibration procedure to ensure accurate monitoring of the participant's dominant eye (which was the only eye tracked). The procedure was performed and repeated until there was less than a maximum error of 0.5 degrees. If it was not possible to meet this criterion,

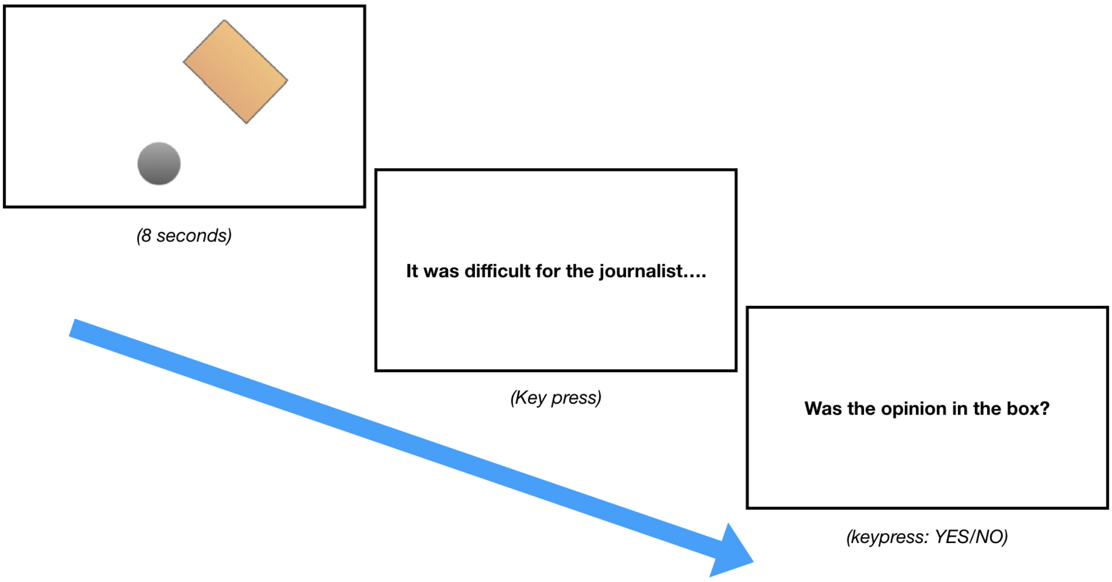


**Figure 3.2:** Example of the progresssion of a trial in experiments 1-3

the experiment was aborted and participants were replaced. Re-calibration was performed after every pause in the Experiment, i.e. twice more. After calibration, participants saw three practice trials before the experiment began. Each trial in the experiment consisted of three phases (see Figure 3.2): First, participants saw an animated video presented on the screen for 8 seconds. The video disappeared and a sentence appeared on the screen. Participants read the sentence and pressed a button on a Cedrus response pad that was in front of them when they had finished reading. The sentence then disappeared and a question appeared on the screen. Participants had to answer this question by pressing either the YES or NO button on the pad (position of YES and NO buttons on either right or left side of the pad was counterbalanced across participants).

### 3.1.6 Analysis & results

#### Analysis of eye-tracking data

Prior to analysis, an intercepts-only regression model was fitted to the data in order to observe the distribution of the residuals. These were not normally distributed (which violates the assumptions of the linear model), and thus a box-cox test (Box

& Cox, 1964) was performed (see Vasishth & Nicenboim, 2016, for more information on this procedure and why it is necessary). The test showed that the reading times measures needed to be transformed using a Lambda value of -0.7, which was used for transforming all eye-tracking measures and regions. Cases in which participants gave an incorrect answer to the comprehension question were also excluded from all analyses. This procedure was followed for all subsequent experiments. Accuracy for comprehension questions in experiment 1 was above 85% in all conditions. I analyzed all data in the experiments using the R statistical programming environment and the LME4 package for regression analysis. To test the predictions, I fitted mixed-effects linear regression models to every measure and every region.

For constructing the statistical models, I followed the recommendations of Barr et al. (2013). First, I tried fitting the largest possible random effects structure granted by the experimental design (in this case, random intercepts and slopes by items and subjects for both independent variables). If the model failed to converge, I reduced the random effects structure step-wise until a converging model was found by first removing the random correlations, then the random intercepts, followed by the interaction effects and the main effects. I used the same maximally converging random effects structure for all dependent measures in every region for the sake of consistency.

All models included trial order as a fixed effect, since it significantly improved the model fit. The models were fitted using an ANOVA-style, sum-contrast coding scheme, which tests for main effects and interactions (unless explicitly stated otherwise). Alpha thresholds for assessing statistical significance for eye-tracking reading data were Bonferroni-corrected, following the recommendations of von der Malsburg and Angele (2017).

The final random effects structure used for every model is shown in Table 3.12. Figures 3.3, 3.4, 3.5 show bar-plots of the results in the adjective, noun, and verb region respectively. Figure 3.6 shows the results of the post-sentence comprehension question response times.

A note on all tables in this chapter reporting the results of regression analyses (as well as in all subsequent chapters): For every variable, the tables report the regression coefficient, followed by the confidence intervals (in brackets) and the t-value. Stars on the t-values depict the statistical significance cut-offs, based on the variable's p-value.

#### Results of eye-tracking, adjective region

No significant main effects or interactions were found in any measure for this region.

#### Results of eye-tracking, target noun region

As predicted, there was a significant main effect of prime-type in all three measures, with shorter reading times in the noun-label vs. no-label conditions. This means that when participants saw the word *opinion* in the video this facilitated reading the same word when it appeared later in the sentence. The experiment was therefore sensitive enough to detect identity priming effects, which suggests that participants were actively integrating the information processed during the video with the information from the sentence.

#### Results of eye-tracking, verb region

No significant main effects or interactions of our manipulated variables were found in any measure for this region.

#### Analysis and results of question response times

Question-response times were time-locked to the appearance of the question on the participant's screen. A box-cox test determined that the response times needed to be log-transformed in order to normalize the residuals of the model. I thus fitted a linear mixed-effects regression model to the log-transformed reaction times. This model was fitted only to correct responses, which were over 92% of all trials.

The results pattern can be seen in Figure 3.6 and the output of the model is summarized in Table 3.4. There was a main effect of question type, showing that participants were significantly slower at answering questions in the noun vs. video

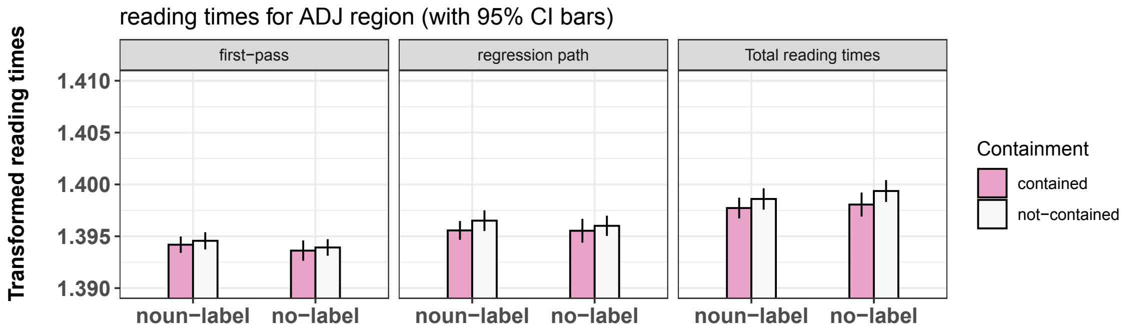


**Figure 3.3:** Summary of results for the ADJ region, Experiment 1

conditions. There was also a main effect of prime type, indicating that participants were faster to answer questions in the noun-label compared to the no-label condition, and a main effect of containment, showing that there was an overall facilitation in the contained vs. not-contained conditions. There were also significant interactions between question type and prime type and containment and prime type, reflecting that the noun-question and video-question conditions had very different response time patterns. The three-way interaction was not significant.

However, these results are not straight-forward to interpret since a response bias was discovered after running the experiment: The correct answer to the question asked was always NO in the not-contained conditions and YES in the contained conditions. It is therefore not possible to tell whether the facilitation effect was caused by the difference in the conditions (contained vs. not-contained) or by the differences in correct answer (YES vs. NO).

The noun-question/no-label was the only exception to this: Here, the correct response was NO in both contained and not-contained levels. Because of this, I re-fitted the statistical model for the question-response times using a treatment contrast coding scheme in order to look at the noun-question/no-label condition exclusively. This type of contrast coding allows for direct comparisons between the condition set as the intercept of the model and the other individual conditions. This model showed no significant difference between contained and not-contained levels of the noun-question/no-label. This model is shown in Table 3.10.

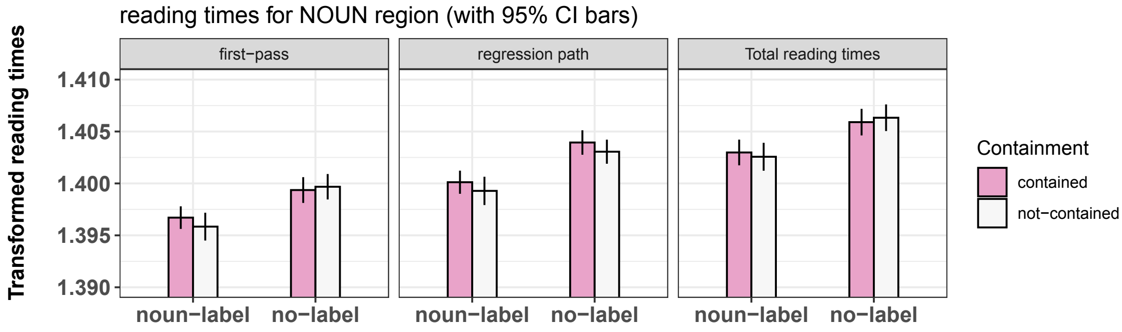


**Figure 3.4:** Summary of results for the TARGET NOUN region, Experiment 1

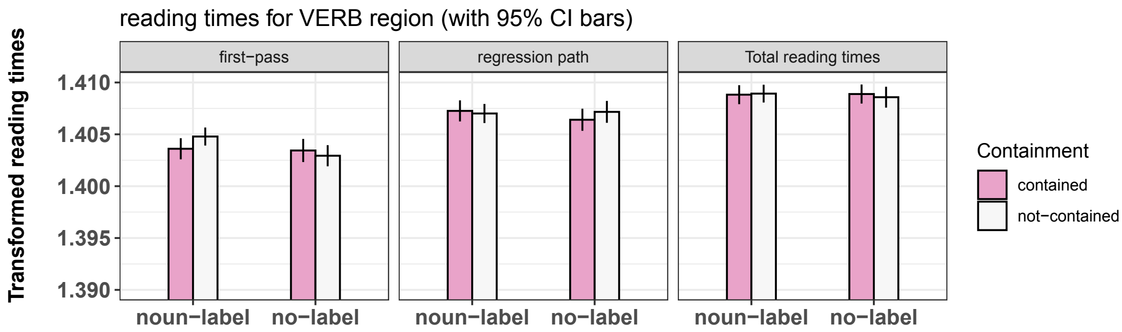


**Figure 3.5:** Summary of results for the VERB region, Experiment 1

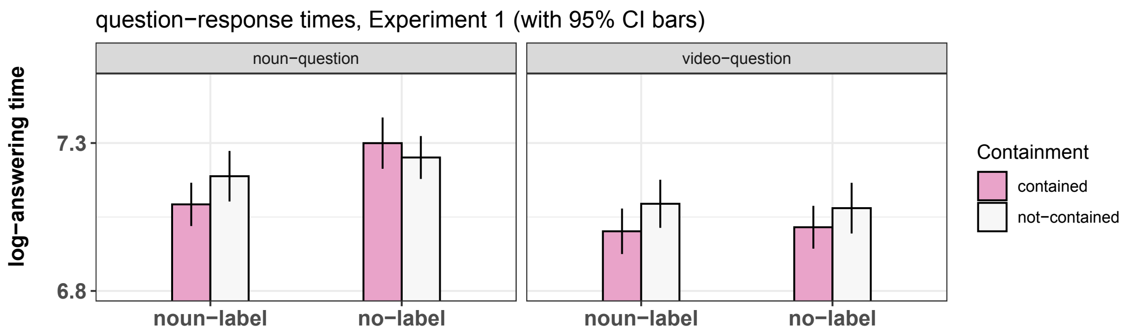


**Figure 3.6:** Summary of results for the question response time, Experiment 1

**Table 3.1:** Regression analysis of reading times in the ADJECTIVE region of Experiment 1

| | *Dependent variable:* | | |
|---|---|---|---|
| | First-pass | Regression path | Total reading times |
| | (1) | (2) | (3) |
| Prime Type | 0.0003 | 0.0001 | −0.0002 |
| | (−0.0001, 0.001) | (−0.0003, 0.001) | (−0.001, 0.0004) |
| | t = 1.444 | t = 0.581 | t = −0.795 |
| Containment | −0.0002 | −0.0003 | −0.0005 |
| | (−0.001, 0.0003) | (−0.001, 0.0001) | (−0.001, 0.00000) |
| | t = −0.744 | t = −1.389 | t = −1.956 |
| Trial Order | −0.00000 | −0.00001 | −0.00003 |
| | (−0.00002, 0.00001) | (−0.00003, 0.00000) | (−0.00004, −0.00001) |
| | t = −0.293 | t = −1.512 | t = −3.105** |
| Interaction | −0.00002 | −0.0001 | 0.0001 |
| | (−0.0004, 0.0004) | (−0.001, 0.0004) | (−0.0004, 0.001) |
| | t = −0.107 | t = −0.494 | t = 0.338 |
| Intercept | 1.394 | 1.397 | 1.400 |
| | (1.393, 1.395) | (1.396, 1.398) | (1.399, 1.401) |
| | t = 3,233.003*** | t = 2,739.009*** | t = 2,625.450*** |
| Observations | 1,180 | 1,180 | 1,180 |
| Log Likelihood | 4,162.329 | 3,965.666 | 3,910.930 |
| Akaike Inf. Crit. | −8,300.659 | −7,907.333 | −7,797.859 |
| Bayesian Inf. Crit. | −8,239.780 | −7,846.453 | −7,736.980 |

*Note:* *p<0.017; **p<0.0033; ***p<0.00033
significance thresholds are bonferroni-corrected (alpha/3), values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table 3.2:** Regression analysis of reading times in the TARGET NOUN region of Experiment 1

| | *Dependent variable:* | | |
|---|---|---|---|
| | First-pass | Regression path | Total reading times |
| | (1) | (2) | (3) |
| Prime Type | −0.002 | −0.002 | −0.002 |
| | (−0.002, −0.001) | (−0.003, −0.001) | (−0.002, −0.001) |
| | t = −5.490*** | t = −6.144*** | t = −5.230*** |
| Containment | 0.0002 | 0.0004 | 0.00003 |
| | (−0.001, 0.001) | (−0.0003, 0.001) | (−0.001, 0.001) |
| | t = 0.430 | t = 1.221 | t = 0.075 |
| Trial Order | −0.00000 | −0.00001 | −0.00004 |
| | (−0.00002, 0.00002) | (−0.00003, 0.00001) | (−0.0001, −0.00002) |
| | t = −0.072 | t = −1.265 | t = −3.687*** |
| Interaction | 0.0003 | −0.00002 | 0.0002 |
| | (−0.0004, 0.001) | (−0.001, 0.001) | (−0.0005, 0.001) |
| | t = 0.800 | t = −0.054 | t = 0.507 |
| Intercept | 1.398 | 1.402 | 1.407 |
| | (1.397, 1.399) | (1.401, 1.403) | (1.405, 1.408) |
| | t = 2,294.932*** | t = 2,344.835*** | t = 2,192.578*** |
| Observations | 1,111 | 1,111 | 1,111 |
| Log Likelihood | 3,559.108 | 3,578.840 | 3,505.177 |
| Akaike Inf. Crit. | −7,094.215 | −7,133.679 | −6,986.355 |
| Bayesian Inf. Crit. | −7,034.059 | −7,073.523 | −6,926.198 |

*Note:* *p<0.017; **p<0.0033; ***p<0.00033
significance thresholds are bonferroni-corrected (alpha/3), values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table 3.3:** Regression analysis of reading times in the VERB region of Experiment 1

| | *Dependent variable:* | | |
|---|---|---|---|
| | First-pass | Regression path | Total reading times |
| | (1) | (2) | (3) |
| Prime Type | 0.001<br>(0.0001, 0.001)<br>t = 2.180 | 0.0002<br>(−0.0003, 0.001)<br>t = 0.760 | 0.0002<br>(−0.0003, 0.001)<br>t = 0.792 |
| Containment | −0.0001<br>(−0.001, 0.0004)<br>t = −0.496 | −0.0001<br>(−0.001, 0.0004)<br>t = −0.459 | 0.0001<br>(−0.0003, 0.001)<br>t = 0.611 |
| Trial Order | −0.00002<br>(−0.00003, −0.00000)<br>t = −2.234 | −0.00001<br>(−0.00003, 0.00000)<br>t = −1.383 | −0.0001<br>(−0.0001, −0.0001)<br>t = −9.256*** |
| Interaction | −0.0004<br>(−0.001, 0.0001)<br>t = −1.683 | 0.0002<br>(−0.0003, 0.001)<br>t = 0.891 | −0.0002<br>(−0.001, 0.0003)<br>t = −0.736 |
| Intercept | 1.405<br>(1.404, 1.406)<br>t = 2,726.656*** | 1.408<br>(1.407, 1.409)<br>t = 2,704.316*** | 1.413<br>(1.412, 1.413)<br>t = 3,084.624*** |
| Observations | 1,148 | 1,148 | 1,148 |
| Log Likelihood | 3,877.586 | 3,866.024 | 4,012.204 |
| Akaike Inf. Crit. | −7,731.172 | −7,708.049 | −8,000.409 |
| Bayesian Inf. Crit. | −7,670.623 | −7,647.499 | −7,939.859 |
| *Note:* | *p<0.017; **p<0.0033; ***p<0.00033<br>significance thresholds are bonferroni-corrected (alpha/3), values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) | | |

### 3.1.7 Discussion

In Experiment 1, no difference in reading times between conditions in the adjective region was found. More importantly, there were no differences in the verb region, the main interest region of the experiment. However, the presence of the effect of priming type in the TARGET NOUN region suggests that the absence of an effect of containment might be interpreted meaningfully: It could be the case that no effect of containment on reading times of the verb was found because

**Table 3.4:** Regression analysis of response-times in Experiment 1

| | *Dependent variable:* |
|---|---|
| | Response times (in log-milliseconds) |
| Containment | −0.021<br>(−0.044, 0.002)<br>t = −1.813 |
| Prime Type | −0.027<br>(−0.050, −0.004)<br>t = −2.325* |
| Question Type | 0.081<br>(0.057, 0.104)<br>t = 6.754*** |
| Trial Order | −0.004<br>(−0.004, −0.003)<br>t = −9.385*** |
| Containment*Prime Type interaction | −0.027<br>(−0.049, −0.004)<br>t = −2.267* |
| Containment*Question Type interaction | 0.012<br>(−0.011, 0.035)<br>t = 1.032 |
| Question Type*Prime Type interaction | −0.038<br>(−0.061, −0.015)<br>t = −3.269** |
| Three-way interaction | −0.016<br>(−0.039, 0.007)<br>t = −1.345 |
| Intercept | 7.327<br>(7.260, 7.394)<br>t = 215.365*** |
| Observations | 1,111 |
| Log Likelihood | −559.837 |
| Akaike Inf. Crit. | 1,145.674 |
| Bayesian Inf. Crit. | 1,210.843 |
| *Note:* | *p<0.05; **p<0.01; ***p<0.001<br>values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) |

the feature of containment is not relevant for the construction of the metaphoric meaning and it is thus ignored during processing, exerting neither facilitation nor interference. This interpretation would be broadly compatible with the Category Inclusion view, which would posit either an interference effect of the literal feature on metaphor comprehension or a null-effect.

However, it might also be possible that no effect was found given the temporal distance between presentation of the visual prime and reading of the metaphorically used verb. Perhaps this distance made it so that, when reading the verb, participants no longer held the activated feature of containment in working memory. It has previously been stated that people struggle holding a large number of meaningful units in working memory (no more than 7, Chen & Cowan, 2005; Miller, 1956). This lays the groundwork for Experiment 2, in which I changed the sentence structure so that the verb could be temporally closer to the video prime.

Results from the post-sentence comprehension questions present an intricate pattern. There was a main effect of question type, with longer response times in the noun-question conditions than in the video-question conditions. There was a main effect of containment, with shorter response times in the contained compared to the not-contained conditions in all but the noun-question/ no-label conditions (as evidenced by the interaction effect between containment and question type).

At first glance, the results seem to support the idea that when the conceptual feature of containment was activated by the verb it generally facilitated responses, resulting in shorter response times in the contained vs. not-contained conditions in all but the noun-question/no-label conditions. This could suggest that the simple feature of containment was activated after the metaphor was understood, but not the more complex mapping between containment and the metaphoric topic (which would have caused an difference in the noun-question/no-label conditions), compatible with Carston's (2010) view on the "lingering" of the literal meaning, but incompatible with the stronger view of Gentner et al. (2001), according to which the pattern of mappings should remain available for further processing.

There is, however, a simpler explanation. As mentioned in the results section, the correct responses were confounded with the conditions contained and not-contained conditions, with contained conditions always requiring a YES response and not-contained conditions a NO response in all but the noun-question/ no-label conditions, were the correct response was NO in both levels of containment. It is therefore likely that it was simply easier for participants to answer YES than to answer NO, explaining the main effect of containment. Additionally, the effect of question type could be due to the fact that questions in the "noun" conditions (which varied according to the target noun in every trial, 33 characters on average) were on average longer than the questions in the "video" conditions (which were always the same, i.e. *Was the ball in the box?*, 30 characters). It is possible that participants just took longer to read the questions in the noun compared to the video conditions and thus took longer to answer the question.

The only comparison not affected by these two issues was that between contained and not-contained levels of the noun-question/no-label condition. For these two levels, the question and correct response remained the same (i.e. NO). There was no significant difference between these two conditions. It's important to note, however, that the YES/NO confound affected only the question response times and not the eye-tracking data, since participants only responded YES or NO to the comprehension question, but no such action was required when reading the metaphoric sentence. I address the issue of the interpretation of question-response times in Experiment 3, where I examine the response patterns to the same questions in the absence of metaphorical verbs. For now, I turn to Experiment 2, where I attempted to replicate the pattern of reading times displayed in Figure 3.5 using sentences with a different syntactic structure.

## 3.2 Experiment 2

The goal of Experiment 2 was to determine the robustness of the results of Experiment 1. First, I altered the sentence structure in order to minimize the temporal distance between prime and verb. I did this because I thought it was likely that participants were not able to use the information extracted from the visual prime to facilitate processing of the metaphoric verb due to working memory constraints. This possibility finds some support in the literature on working memory, where it has been noted that people have a relatively low average number of sequentially presented meaningful units that they can remember (somewhere between 3 and 7, Chen & Cowan, 2005; Miller, 1956). I also increased the number of participants, from 48 to 64, to obtain higher statistical power. I did this following a power analysis via simulation using the R package SimR (Green & MacLeod, 2016). For the power analysis, I took the model of the total reading times for the verb region as starting point. The simulations suggested that with 64 participants the experiment would have over 80% power to detect a main effect of containment on total reading times of the verb region, assuming a true effect size of Cohen's d= 0.15, i.e. somewhat smaller than the rule of thumb for a "small" effect size (Sawilowsky, 2009). By doing this I aimed to either detect an effect that was not detected in the previous experiment, or to replicate the reading-time pattern and null-effect of Containment in the verb region of Experiment 1 with higher statistical power.

### 3.2.1 Participants

Sixty-four native speakers of German (ages 18–31) with normal or corrected-to-normal vision were recruited and tested at the Humboldt-Universität zu Berlin. None of them participated in Experiment 1. They received 8 euros as compensation upon giving their informed consent. This study was covered by the ethics vote of the psycholinguistics lab of the Humboldt-Universität zu Berlin.

### 3.2.2 Materials, design and procedure

The materials, design, and procedure were identical to those in Experiment 1, except for the syntactic structure of the critical sentence, which now displayed a leftward movement of the subject clause. This allowed for the verb to appear as the fourth word in the sentence, making it temporally closer to the video-prime than it was in

Experiment 1. The structure of the sentences was as follows:

(3) *Dass seine / Meinung* TARGET NOUN */ umgittert* VERB */ wurde nach dem Regimewechsel, war / schwierig* ADJ */ für den Redakteur.*

*"That his / opinion* TARGET NOUN */ fenced-in* VERB */ was after the change in regime, was /difficult* ADJ*/ for the journalist"*

"The fact that his opinion was fenced-in after the change in regime was difficult for the journalist."

### 3.2.3 Predictions

The predictions were motivated by the results of Experiment 1: If the absence of an effect of containment on the verb region was due to the temporal distance between verb and video, moving the verb closer to the video should correct this. Specifically, if priming physical containment facilitates processing of verbs of spatial containment used metaphorically, this should be reflected in shorter reading times in the contained vs. not-contained conditions in the VERB region.

With regards to the question-answering times: The overall facilitation effect of contained versus not-contained in experiment 2 was confounded with the type of response ("YES" for contained and "NO" for not-contained) in all but one relevant comparison: The noun-question/no-label conditions. We did not find a significant difference between these two conditions. In Experiment 2 I hoped to replicate the question-answering pattern in general, and the results of the noun-question/no-label conditions in particular.

### 3.2.4 Results

**Eye-tracking**

Results for all regions and measures are shown in Figures 3.7, 3.8, 3.9 and 3.10. The output of the statistical models can be seen in Tables 3.5, 3.6, 3.7 and 3.8.

**Adjective** No significant effects of containment or of prime type were found in this region, replicating the pattern found in Experiment 1.

**Target noun** The main effect of prime type was replicated on all measures, with the noun-label conditions showing overall shorter reading times than the no-label conditions. This shows that participants were able to successfully relate the video to the sentence, leading to a reliable priming effect.

**Verb** The analysis failed to show an effect of containment on any measure, as was the case in Experiment 1. There was also no effect of prime type and no significant interaction of containment and prime type.

#### Question-response times

Question-response times were analyzed in the same way as in Experiment 1. As can be seen in Figure 3.10, the results are very similar to those of Experiment 1. All previous findings were replicated with the exception of the main effect of containment: There was a main effect of question type and of prime type. There was an interaction between containment and question type and an interaction between question type and prime type. This model can be seen in Table 3.8. As in the previous experiment, I re-fitted the model using a treatment-contrast scheme in order to directly compare contained and not-contained levels of the noun-question/no-label condition. This model showed no significant difference between these conditions, replicating the result found in Experiment 1 (see Table 3.10).

### 3.2.5 Discussion

In Experiment 2 I tried to facilitate the interaction between video prime and metaphoric verb by increasing statistical power and decreasing the temporal distance between verb and video. The experiment still failed to show an effect of containment in the verb region. Besides this, the effect of prime type was replicated on all measures in the target noun region: Seeing the word *opinion* written on the ball

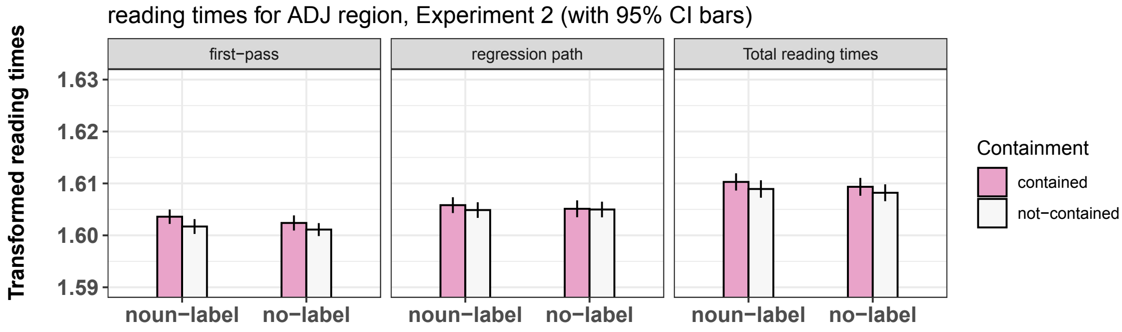


**Figure 3.7:** Summary of results for the ADJ region, Experiment 2

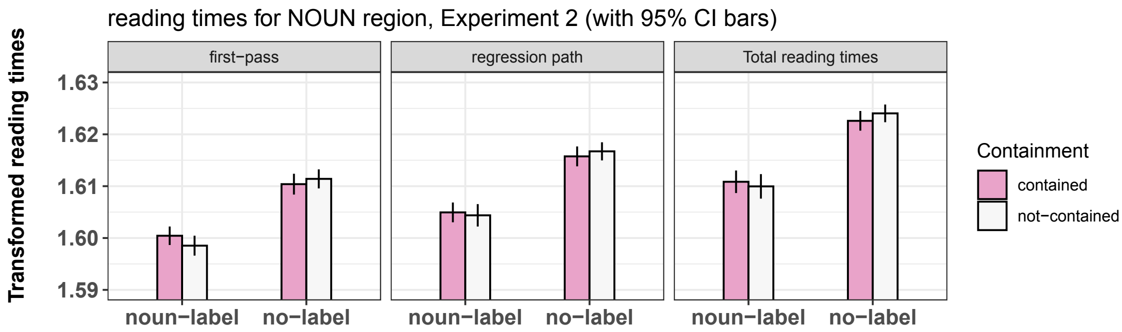


**Figure 3.8:** Summary of results for the TARGET NOUN region, Experiment 2

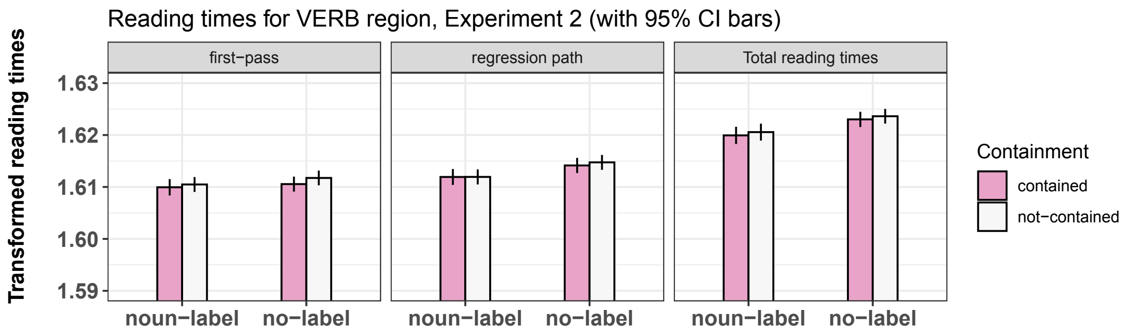


**Figure 3.9:** Summary of results for the VERB region, Experiment 2

**Table 3.5:** Regression analysis of reading times in the ADJECTIVE region of Experiment 2

| | *Dependent variable:* | | |
|---|---|---|---|
| | First-pass | Regression path | Total reading times |
| | (1) | (2) | (3) |
| Prime Type | 0.0005 | 0.0002 | 0.0005 |
| | (−0.0002, 0.001) | (−0.001, 0.001) | (−0.0004, 0.001) |
| | t = 1.356 | t = 0.445 | t = 1.029 |
| Containment | 0.001 | 0.0002 | 0.001 |
| | (0.0001, 0.001) | (−0.0005, 0.001) | (−0.0002, 0.001) |
| | t = 2.329 | t = 0.660 | t = 1.389 |
| Trial Order | −0.00001 | −0.00002 | −0.0001 |
| | (−0.00003, 0.00001) | (−0.00004, 0.00001) | (−0.0001, −0.00003) |
| | t = −0.619 | t = −1.480 | t = −3.907*** |
| Interaction | 0.0001 | 0.0002 | 0.00001 |
| | (−0.001, 0.001) | (−0.001, 0.001) | (−0.001, 0.001) |
| | t = 0.433 | t = 0.479 | t = 0.026 |
| Intercept | 1.603 | 1.606 | 1.612 |
| | (1.601, 1.604) | (1.605, 1.608) | (1.610, 1.614) |
| | t = 2,303.281*** | t = 2,069.854*** | t = 1,926.601*** |
| Observations | 1,634 | 1,634 | 1,634 |
| Log Likelihood | 4,709.178 | 4,530.882 | 4,406.515 |
| Akaike Inf. Crit. | −9,394.356 | −9,037.763 | −8,789.029 |
| Bayesian Inf. Crit. | −9,329.571 | −8,972.978 | −8,724.244 |
| *Note:* | *p<0.017; **p<0.0033; ***p<0.00033 | | |

significance thresholds are bonferroni-corrected (alpha/3), values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table 3.6:** Regression analysis of reading times in the TARGET NOUN region of Experiment 2

| | *Dependent variable:* | | |
|---|---|---|---|
| | First-pass | Regression path | Total reading times |
| | (1) | (2) | (3) |
| Prime Type | −0.006<br>(−0.007, −0.004)<br>t = −9.825*** | −0.006<br>(−0.007, −0.005)<br>t = −10.309*** | −0.006<br>(−0.007, −0.005)<br>t = −12.163*** |
| Containment | 0.0002<br>(−0.001, 0.001)<br>t = 0.465 | −0.0002<br>(−0.001, 0.001)<br>t = −0.334 | −0.0002<br>(−0.001, 0.001)<br>t = −0.469 |
| Trial Order | −0.00002<br>(−0.0001, 0.00000)<br>t = −1.618 | −0.00005<br>(−0.0001, −0.00002)<br>t = −3.415** | −0.0001<br>(−0.0001, −0.00004)<br>t = −4.296*** |
| Interaction | 0.001<br>(−0.001, 0.002)<br>t = 1.021 | 0.0002<br>(−0.001, 0.002)<br>t = 0.294 | 0.0004<br>(−0.001, 0.002)<br>t = 0.575 |
| Intercept | 1.607<br>(1.605, 1.608)<br>t = 1,744.954*** | 1.613<br>(1.611, 1.615)<br>t = 1,728.147*** | 1.621<br>(1.619, 1.623)<br>t = 1,625.652*** |
| Observations | 1,491 | 1,491 | 1,491 |
| Log Likelihood | 3,908.996 | 3,884.676 | 3,796.362 |
| Akaike Inf. Crit. | −7,793.993 | −7,745.352 | −7,568.724 |
| Bayesian Inf. Crit. | −7,730.306 | −7,681.666 | −7,505.037 |

*Note:* *p<0.017; **p<0.0033; ***p<0.00033
significance thresholds are bonferroni-corrected (alpha/3), values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

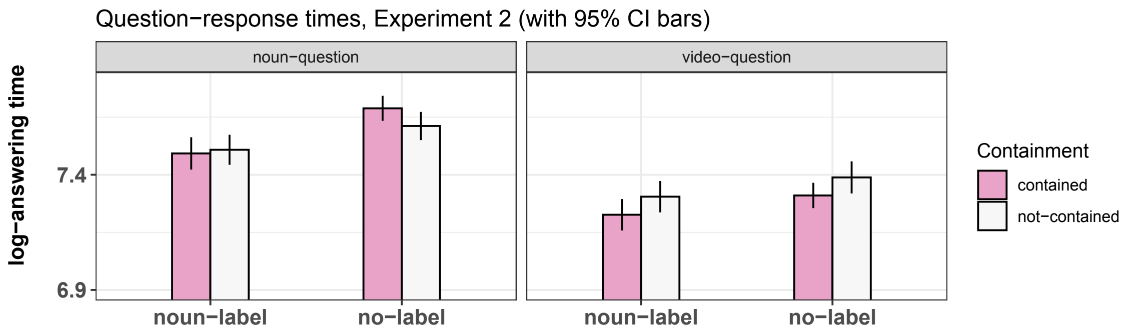


**Figure 3.10:** Summary of results for the question response time, Experiment 2

**Table 3.7:** Regression analysis of reading times in the VERB region of Experiment 2

| | *Dependent variable:* | | |
|---|---|---|---|
| | First-pass | Regression path | Total reading times |
| | (1) | (2) | (3) |
| Prime Type | −0.0004 | −0.001 | −0.001 |
| | (−0.001, 0.0003) | (−0.002, −0.0004) | (−0.002, −0.001) |
| | t = −1.165 | t = −3.061** | t = −3.779*** |
| Containment | −0.0005 | −0.0002 | −0.0004 |
| | (−0.001, 0.0003) | (−0.001, 0.0005) | (−0.001, 0.0003) |
| | t = −1.153 | t = −0.550 | t = −1.129 |
| Trial Order | −0.00004 | −0.00005 | −0.0001 |
| | (−0.0001, −0.00002) | (−0.0001, −0.00002) | (−0.0001, −0.0001) |
| | t = −3.886*** | t = −4.105*** | t = −8.117*** |
| Interaction | 0.0001 | 0.0001 | −0.0002 |
| | (−0.001, 0.001) | (−0.001, 0.001) | (−0.001, 0.001) |
| | t = 0.270 | t = 0.191 | t = −0.370 |
| Intercept | 1.613 | 1.616 | 1.627 |
| | (1.612, 1.615) | (1.614, 1.617) | (1.626, 1.629) |
| | t = 2,206.844*** | t = 2,209.841*** | t = 2,151.451*** |
| Observations | 1,566 | 1,566 | 1,566 |
| Log Likelihood | 4,459.908 | 4,456.508 | 4,402.584 |
| Akaike Inf. Crit. | −8,895.816 | −8,889.016 | −8,781.167 |
| Bayesian Inf. Crit. | −8,831.541 | −8,824.741 | −8,716.892 |

*Note:* *p<0.017; **p<0.0033; ***p<0.00033
significance thresholds are bonferroni-corrected (alpha/3), values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table 3.8:** Regression analysis of response-times in Experiment 2

| | *Dependent variable:* |
|---|---|
| | Response times (in log-milliseconds) |
| Containment | −0.017<br>(−0.034, 0.001)<br>t = −1.817 |
| Prime Type | −0.060<br>(−0.078, −0.042)<br>t = −6.551*** |
| Question Type | 0.134<br>(0.116, 0.152)<br>t = 14.741*** |
| Trial Order | −0.004<br>(−0.004, −0.003)<br>t = −13.198*** |
| Containment*Prime Type interaction | −0.018<br>(−0.036, −0.00003)<br>t = −1.963* |
| Containment*Question Type interaction | 0.029<br>(0.011, 0.047)<br>t = 3.178** |
| Question Type*Prime Type interaction | −0.013<br>(−0.031, 0.005)<br>t = −1.439 |
| Three-way interaction | −0.007<br>(−0.025, 0.011)<br>t = −0.722 |
| Intercept | 7.663<br>(7.602, 7.724)<br>t = 246.117*** |
| Observations | 1,491 |
| Log Likelihood | −613.253 |
| Akaike Inf. Crit. | 1,252.506 |
| Bayesian Inf. Crit. | 1,321.499 |
| *Note:* | *p<0.05; **p<0.01; ***p<0.001<br>values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) |

in the video facilitated reading times of that same word once it appeared in the sentence. This confirms that participants were able to use the information presented in the video to ease processing of the target noun, and that they were nevertheless unable to use the feature of "containment" presented in the video to speed up (or slow-down) reading times in the verb region. This suggests that during processing of the metaphoric verb, participants largely ignored the feature of physical containment, seeing as it neither interfered with nor facilitated processing. This is consistent with a Category Inclusion view of metaphor comprehension that states that literal features are not initially activated if they are not necessary for the construction of the appropriate ad hoc category during metaphor processing.

However, it could also be the case that the lack of effects in the verb region is caused by inadequate materials: Activating the feature of spatial containment could indeed facilitate or hinder processing, but the participants were not able to activate this feature from these specific videos. It is thus necessary to assess whether these videos could modulate processing in an environment in which they would be expected to do so reliably, namely when the verbs are processed in their encoded, literal meaning only. If the videos facilitate access to the literal meaning of the verbs, the current interpretation of the results of Experiments 1 and 2 becomes more plausible. I address this issue in Experiment 4.

The results of the question response task broadly replicated the findings of Experiment 1. It was easier for participants to answer the question in the contained vs. not-contained levels of the video-question conditions. In the noun-question conditions, there was an effect of prime type, with the no-label conditions showing slower response times than the noun-label conditions.

The noun-question/no-label conditions did not show a significant difference between contained and not-contained levels, just as in Experiment 1. This finding is important because the noun-question / no-label conditions were the only ones without a confound between condition and correct answer. Furthermore, there was an effect of prime type in the noun-question conditions, with the "noun-label" conditions showing longer response times than the "noun-label" conditions.

As mentioned in the discussion of Experiment 1, these results could be interpreted as meaning that when reading the sentence, the conceptual feature of containment is activated, facilitating responses in the contained vs. not-contained conditions and/or interfering with the responses in the noun-question/no-label conditions.

This interpretation, however, is contingent upon the assumption that the response patterns were caused by the interaction of processing video and metaphor and not by the YES/NO response confound or by other external factors. I sought to test this assumption in Experiment 3.

## 3.3 Experiment 3

Question-response times in Experiments 1 and 2 show an overall facilitation effect for contained vs not-contained conditions, except for the noun-question/no-label conditions, which showed no difference between contained and not-contained levels. In Experiment 3, I set out to test whether these results were caused by the interaction of video-prime, metaphor and question, or whether they are independent of the presence of the metaphoric sentence. To do this, I ran a version of Experiment 2 in which the sentences read by participants did not contain any metaphors whatsoever: If the same pattern of results as in the previous two experiments is visible, it would suggest that the results are not related to the processing of verbal metaphors. Since the reading patterns of these sentences are irrelevant for this question and only the question-response times were of interest, Experiment 3 was not run as an eye-tracking study. Instead, it was implemented as a self-paced reading reaction time task: Participants first watched the video-prime and then read the (non-metaphoric) sentence. When they were done reading, they pushed a button in front of them and were presented with the comprehension questions, which they answered by pushing either a YES or NO button. Only the response times to the comprehension questions were measured, time-locked to the appearance of the question on the screen, as was the case in Experiments 1 and 2.

### 3.3.1 Participants

Sixty-four native speakers of German (ages 18–31) with normal or corrected-to-normal vision were recruited and tested at the Humboldt-Universität zu Berlin. None of them participated in Experiments 1, or 2. They received 8 euros as compensation upon giving their informed consent. This study was covered by the ethics vote of the psycholinguistics lab of the Humboldt-Universität zu Berlin.

### 3.3.2 Materials and design

To construct the materials in Experiment 3, I modified the sentences from Experiment 2 by replacing the verb with a non-metaphorical verb (e.g. *ignored* instead of *fenced-in*) that did not have the feature of spatial containment as part of its literal meaning, as presented in (4):

(4) *"Dass seine Meinung ignoriert wurde nach dem Regimewechsel, war für den Redakteur schwierig"*

"The fact that his opinion was ignored after the change in regime was difficult for the journalist"

The design was identical to that of the previous experiments, with the factors containment, question type and prime type. The experiment was programmed using the open source software Open Sesame and was run on a PC. The only dependent measure in this experiment was question-response time.

### 3.3.3 Procedure

Participants were instructed to wear noise-reducing headphones throughout the experiment to avoid being distracted by the other participants. Each trial consisted of three phases: First, participants saw the same animated video presented in Experiments 1 and 2. They then read a sentence and pressed the space bar on the keyboards that was in front of them (which was different to the procedure of Experiments 1 and 2, were they pressed a key on a Cedrus response pad, not a computer keyboard). The sentence then disappeared and a question appeared on the screen. They had to answer this question by pressing either the letter F or J (whereas in Experiments 1 and 2 they pressed either the left or right key on the response pad), which were counterbalanced across participants to stand for either YES or NO.

### 3.3.4 Predictions

The predictions for Experiment 3 were derived from the results of Experiments 1 and 2: If the same pattern of results is found in Experiment 3 as the one observed in the previous two Experiments, it would suggest that the results of the question-response times of Experiments 1 and 2 were not a consequence of the way participants processed the metaphoric sentences. If a different pattern than this is found, it would suggest that the results found in Experiments 1 and 2 were (at least partially) caused by the way participants processed the verbal metaphors. In this sense, Experiment 3 serves as a baseline against which the results of the question-response times of Experiments 1 and 2 can be interpreted.

Of particular interest is again the noun-question/no-label condition given that, as in the previous two Experiments, this was the only condition were both levels of the factor containment (contained vs. not-contained) had the same correct response (i.e. NO).

### 3.3.5 Results

I fitted a linear mixed effects regression model to the log-transformed reaction times. Main effects of containment, prime type and question type were found. Additionally, significant interactions of containment and question type, containment and prime type, question type and prime type, and question type, prime type and containment were found. The results are shown in Figure 3.11 and the model details are given in Table 3.9.

Re-fitting the model with treatment contrasts, as I did for the previous experiments, showed a significant difference between contained and not-contained levels of the noun-question/no-label conditions, with the contained condition showing

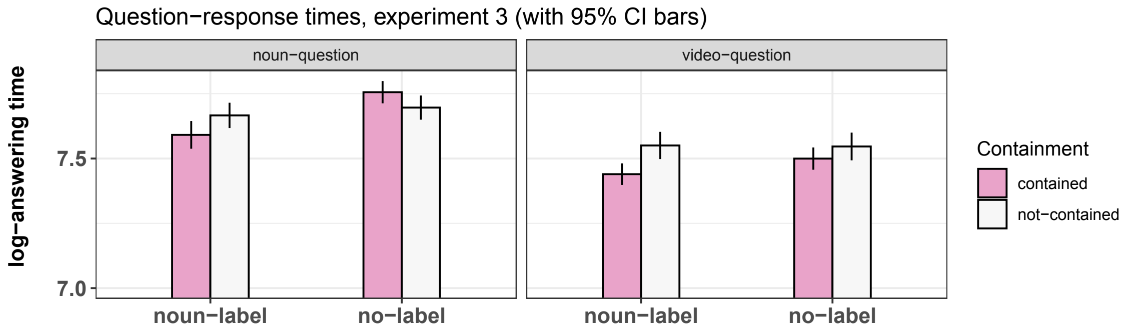


**Figure 3.11:** Summary of results for the question response time, Experiment 3

significantly shorter responses than the not-contained condition. The details of this model are shown in Table 3.10.

### 3.3.6 Discussion

The pattern of results is very similar to that found in Experiments 1 and 2. This suggests that the response times founds in those experiments were modulated by factors other than the metaphorical verb, since there was no metaphorical verb in Experiment 3. This confirms the simpler explanation for the result pattern of Experiments 1 and 2, namely that the results likely follow from a general response bias (Easier to answer YES than NO and easier to answer to shorter than to longer questions), and are not a product of metaphoric interpretation.

However, the results of the noun-question/no-label condition require further explanation. In Experiment 3, both contained and not-contained levels of the noun-question/no-label condition were significantly different from one another, whereas in Experiments 1 and 2, no significant difference between these levels was found. It is thus likely that this difference between experiments is the only one that might be related to the presence of the metaphorical sentences in Experiments 1 and 2: If in the absence of a metaphor there are shorter response times in the not-contained compared to the contained level of the noun-question/no-label condition (Experiment 3, i.e. our baseline result), then the lack of a difference between conditions in the presence of a metaphor (Experiments 1 and 2) could be interpreted as a facilitation effect of the contained compared to the not-contained condition.

**Table 3.9:** Regression analysis of response-times in Experiment 3

| | *Dependent variable:* |
|---|---|
| | Response times (in log-milliseconds) |
| Containment | −0.043<br>(−0.081, −0.006)<br>t = −2.281* |
| Prime Type | 0.168<br>(0.134, 0.203)<br>t = 9.631*** |
| Question Type | −0.062<br>(−0.094, −0.031)<br>t = −3.881*** |
| Trial Order | −0.001<br>(−0.003, 0.001)<br>t = −1.307 |
| Containment*Prime Type interaction | 0.071<br>(0.008, 0.133)<br>t = 2.219* |
| Containment*Question Type interaction | −0.099<br>(−0.165, −0.033)<br>t = −2.936** |
| Question Type*Prime Type interaction | −0.067<br>(−0.130, −0.005)<br>t = −2.105* |
| Three-way interaction | −0.069<br>(−0.194, 0.056)<br>t = −1.088 |
| Intercept | 7.612<br>(7.580, 7.644)<br>t = 466.819*** |
| Observations | 2,113 |
| Log Likelihood | −887.236 |
| Akaike Inf. Crit. | 1,822.472 |
| Bayesian Inf. Crit. | 1,958.213 |
| *Note:* | *p<0.05; **p<0.01; ***p<0.001<br>values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) |

**Table 3.10:** Model fitted with treatment-contrast coding for response times of Experiments 1-3. 'Containment' shows effect in noun-question/no-label conditions only

| | *Dependent variable:* | | |
|---|---|---|---|
| | Response times per Experiment (in log-milliseconds) | | |
| | (1) | (2) | (3) |
| Containment | −0.066 | −0.074 | −0.064 |
| | (−0.157, 0.024) | (−0.154, 0.006) | (−0.121, −0.007) |
| | t = −1.436 | t = −1.803 | t = −2.185* |
| Intercept | 7.506 | 7.913 | 7.947 |
| | (7.416, 7.596) | (7.831, 7.995) | (7.880, 8.013) |
| | t = 162.766*** | t = 188.551*** | t = 235.400*** |
| Observations | 1,111 | 1,491 | 2,113 |
| Log Likelihood | −559.837 | −602.232 | −549.023 |
| Akaike Inf. Crit. | 1,145.674 | 1,256.464 | 1,150.046 |
| Bayesian Inf. Crit. | 1,210.843 | 1,394.452 | 1,297.098 |

*Note:* *p<0.05; **p<0.01; ***p<0.001
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

This would suggest that in Experiments 1 and 2 the feature of containment might have made it easier to answer the questions in the noun-question/no-label condition, which would be in line with the predictions made by Implicit Comparison views. However, this is very tenuous evidence at best (especially when considering that the regression coefficients in all three experiments are very similar, see Table 3.10), and would need to be confirmed by a follow-up experiment in which the absence and presence of a metaphoric verb is directly manipulated.

This interpretation, as well as the interpretation of the results of the gaze record of Experiments 1 and 2, relies on the assumption that participants can indeed derive the conceptual feature of containment from the prime videos and that this feature interacts with the way the verbs are processed. Experiment 4 directly addresses this issue.

## 3.4 Experiment 4

This experiment dealt with the question of whether or not the videos used in Experiments 1-3 can activate a mental representation of containment that leads participants to process verbs of physical containment more readily than when they first see an unrelated video.

### 3.4.1 Materials and design

Experiment 4 was a web-based lexical decision task in which participants saw the same video clips from Experiments 1-3 as primes and then read the same verbs from Experiments 1 and 2, which were presented here without context. The experiment thus only had the factor containment with the levels contained and not-contained.

### 3.4.2 Procedure

A sample of 259 German native speakers (120 female; ages 18-31) was recruited online via the platform “clickworker”. The experiment was designed and run using an instance of the IBEX farm (Drummond, 2013) coupled with the Penncontroller extension (Zehr & Schwarz, 2018), which allows for a simple integration of video and linguistic stimuli. On each trial, participants first saw a video prime and then a target word in the middle of the screen, and had a total of 5 seconds to decide whether the word was a real word by either pressing F (“not a real word”) or J (“real word”). After one practice item, participants were presented with six experimental trials (two critical, four fillers). There was a one second pause in-between trials. Figure 3.12 shows the progression of a single trial in the experiment. The study was cleared by the data protection office of the Humboldt-Universität zu Berlin.

### 3.4.3 Predictions

If the video in the “contained” condition is not capable of eliciting a mental representation of “containment” that can aid lexical recognition of verbs of physical containment, there should be no difference in reaction times between conditions.

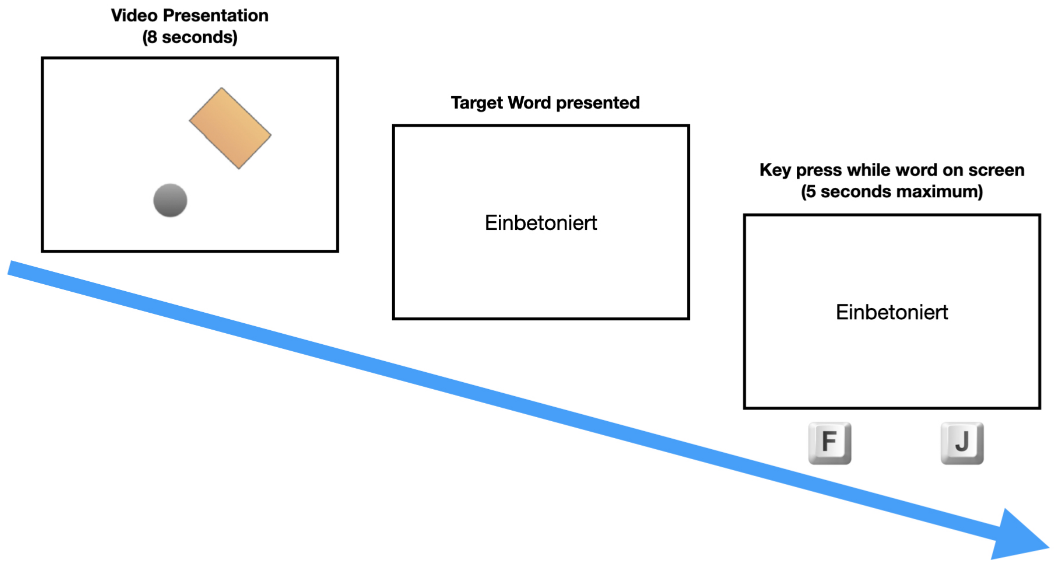


**Figure 3.12:** Example of the progresssion of a trial in Experiment 4

If, on the other hand, the video in the "contained" condition is indeed capable of eliciting a mental representation of "containment" that can ease lexical recognition of verbs of physical containment, shorter reaction times in the contained condition should appear compared to the not-contained condition.

### 3.4.4 Analysis and results

Prior to the analysis, participants who got less than 4/6 correct responses were excluded (n=9), leaving the total number of participants at 250. Reaction times were log-transformed following the results of a box-cox test (Box & Cox, 1964). A linear mixed effects model was then fitted to the data. The results showed a significant difference between the two conditions, with the contained condition displaying shorter reaction times compared to the not-contained condition. The effect size had a value of Cohen's d = 0.21 (i.e. a "small" effect size according to Cohen (1992)). The results are presented in Figure 3.13 and the model summary in Table 3.11.

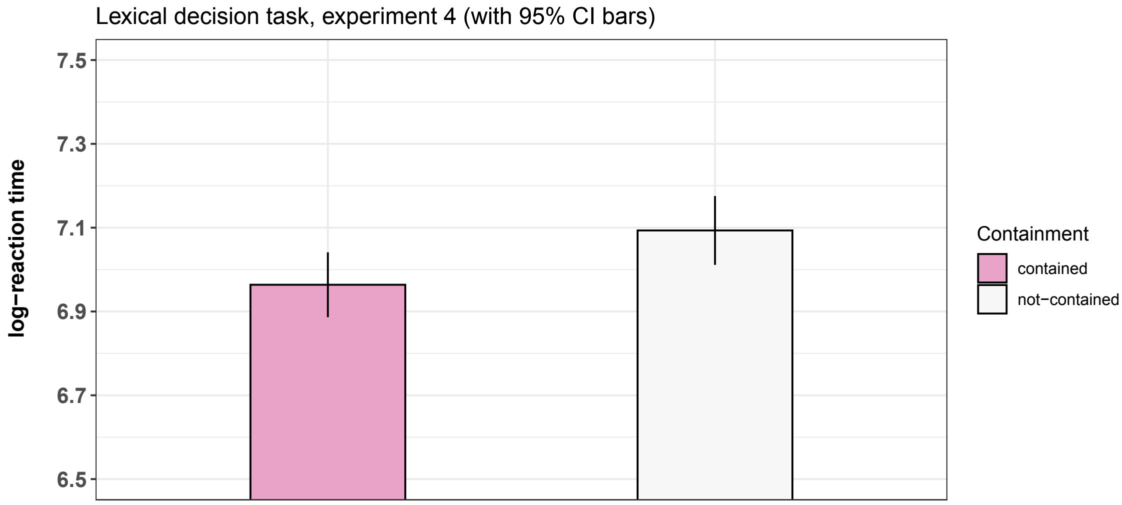


**Figure 3.13:** Summary of results for the lexical decision task, Experiment 4

**Table 3.11:** Regression analysis of response-times in Experiment 4

| | *Dependent variable:* |
|---|---|
| | Response times (in log-milliseconds) |
| Containment | 0.133 |
| | (0.046, 0.221) |
| | t = 2.993** |
| Intercept | 6.974 |
| | (6.885, 7.064) |
| | t = 153.213*** |
| Observations | 465 |
| Log Likelihood | −415.286 |
| Akaike Inf. Crit. | 840.571 |
| Bayesian Inf. Crit. | 861.282 |
| *Note:* | *p<0.05; **p<0.01; ***p<0.001 values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) |

**Table 3.12:** Random effects structure for models in Experiments 1-4

| | Models of Eye-tracking data | Models of forced choice data |
|---|---|---|
| **Exp1** | (0 + prime type * containment \|\|item) + (0 + prime type * containment \|\| subject) | (1 + type \|\| subject) + (0 + type \| item) |
| **Exp2** | (0 + prime type * containment \|\| item)+ (0 + prime type * containment \|\| subject) | (1 + type \|\| subject) + (0 + type \| item) |
| **Exp3** | | (0 + question type * prime type * containment \|\| item) + (0 + containment * question type * prime type \|\|subject) |
| **Exp4** | | (1\|subject) + (1\|item) |

### 3.4.5 Discussion

Experiment 4 showed that the video-clip primes used in Experiments 1-3 facilitated the retrieval of the encoded, literal meaning of different verbs of physical containment. This finding suggests that participants were able to derive the conceptual feature of physical containment from the videos in the contained conditions, since this is the key feature assumed to be shared by video and verbs.

## 3.5 Interim Conclusion

As outlined in chapter two, theories of metaphor processing make different predictions regarding the role of conceptual features related only to the literal meaning during and immediately after processing of novel metaphors. Category Inclusion views believe that these literal features should not play a role during processing and might even hinder comprehension (McGlone & Manfredi, 2001). Furthermore, they should be rapidly suppressed after the metaphor has been understood (Gernsbacher et al., 2001; Rubio Fernandez, 2007). Implicit Comparison views, instead, claim that features related to the literal meaning of a metaphor are initially active. This is caused by an alignment stage in which encoded meanings are fully retrieved prior to the projection of inferences (Gentner & Bowdle, 2008). This means that literal features should facilitate early stages of processing, as shown by Weiland et al. (2014), and can remain active after comprehension, easing understanding of subsequent, related novel or conventional metaphors (Thibodeau & Durgin, 2008).

In the experiments presented in this chapter, I looked at how priming the conceptual feature of spatial containment would interact with the processing of verbal metaphors in which physical containment is a crucial part of the literal meaning but (arguably) not of the metaphoric interpretation. The results of two eye-tracking experiments showed that the videos neither facilitated nor hindered processing of the verbs used (e.g. *fenced-in*), regardless of whether the verb appeared early on or late in the sentence (Experiments 1 and 2). This absence of an effect was accompanied by a reliable priming effect of the target noun that appeared in both video and sentence, suggesting that participants were actively integrating the input of the video with the input of the sentence. Furthermore, Experiment 4 showed that the videos did elicit a priming effect on those same verbs in a de-contextualized lexical decision task, in as much as the contained video elicited shorter response times to the subsequent verbs than did the not-contained video.

Data from the question-response times showed that participants were overall faster answering questions in the contained vs. not-contained conditions. They were also overall slower to answer questions about the interaction between video and sentence (*Was the opinion in the box?*) than about just the video. Since these effects were present in both the experiments with a metaphoric verb (Experiments 1 and 2) and our baseline experiment without a metaphoric verb (Experiment 3) they do not tell us much about how the metaphors interacted with video and question type during processing.

However, in the absence of a metaphor (Experiment 3), participants were significantly faster at correctly answering the question in the noun-question/no-label not-contained condition (*Was the opinion in the box?*, when there was no word written on the ball and the ball bounced freely) compared to the noun-question/no-label contained condition (*Was the opinion in the box?*), when there was no word written on the ball and the ball was trapped by the box). In Experiments 1 and 2, there was no difference between these conditions. This suggests that, in the presence of a metaphor, there could be a facilitation effect of the contained compared to the not-contained noun-question/no-label conditions, which might mean that the metaphor itself activated the feature of spatial containment which later facilitated response times to the post-sentence questions. However, the evidence for this is very tenuous since the overall question-response pattern in all three experiments was similar.

I interpret the data as showing that the feature of physical containment is ignored during comprehension of verbal metaphors of containment and neither facilitates nor hinders processing. Failing to find a significant difference between conditions is not equivalent to finding that there is no difference between them. However, given the results of Experiment 4 and the fact that in Experiments 1 and 2 there was a significant effect of prime type (showing that some aspects of the prime were indeed integrated with the sentence), I believe that the absence of an effect of containment in Experiments 1 and 2 can be interpreted as meaningful.

I see this as being in line with a metaphor processing view that does not ascribe an important role to literal features of the metaphoric vehicle during initial stages of processing. Such is the case of Category Inclusion views (Glucksberg, 2001; Sperber & Wilson, 2008), which claim that the meaning of the vehicle is quickly modulated given the dimensions provided by the topic. In this process, features of the literal meaning that are not compatible with the dimensions provided by the topic do not need to be activated. However, pre-activating these features does not interfere with the lexical modulation of the metaphoric vehicle either.

It could be that metaphor processing varies according to syntactic class such that nominal metaphors are processed differently than verbal metaphors. This would mean that nominal metaphors could be understood via Implicit Comparison (following Gentner & Bowdle, 2008) and verbal metaphors via lexical modulation (Torreano et al., 2005). This view has previously been stated by Cardillo et al. (2010) and Schmidt et al. (2010). However, evidence from a neuroimaging study suggests that on a neural level, the distinction between verbal and nominal metaphors might is not a meaningful one. Cardillo et al. (2012) investigated processing of both nominal and verbal metaphors using functional magnetic resonance imaging technology: In the experiment, participants read nominal (*The reception was an icy swim*) and verbal (*The flowers purred in the sunlight*) metaphors while their brains' blood oxygenation levels were measured. The results showed that the activated regions associated with processing each type of metaphor greatly overlapped, with no significant differences between nominal and verbal metaphors. This suggests that the neural processes associated with processing both of these types of metaphors might not differ, which points to the underlying cognitive mechanisms likely being the same. With this reasoning in mind, it is likely that the present results could generalize beyond the case of verbal metaphors.

In terms of how these results relate to the literature on the interaction between language and the visual world it is possible to draw the following conclusions: Guerra and Knoeferle (2014) found a facilitation effect of visual primes of distance on processing of semantic similarity. They argued that this was indicative of an abstract co-indexing link between distance and similarity. Experiments 1 and 2 of the current chapter failed to find such a link between videos of containment and adjectives of difficulty. It could be the case that these co-indexing links are constructed and stored in memory via repeated, conventional use: Perhaps speaking of semantic similarity in terms of distance is a more common occurrence than speaking of difficulty in terms of containment, leading to facilitation effects in the former but not in the latter case.

In a production study, Sato et al. (2015) found a priming effect of metaphors of difficulty after participants saw images of physical containment, an effect which we failed to find in the present language comprehension study. This difference in results could be explained by a difference in conventionality of the types of metaphors used: Sato et al. (2015) counted the production of spatial prepositions such as *in* and *out* (e.g. *Bobbie fell in love working in the potato factory*) and of idiomatic expressions (*Nick said time is full of shit*) as instances of a containment-as-difficulty metaphor. These types of conventional, "fossilized" metaphoric expressions are likely to be processed differently than novel metaphors (Bowdle & Gentner, 2005; Keysar et al.,

2000) making the results difficult to compare, given that the materials in our study were all novel verbal metaphors (It is not clear whether participants in the study by Sato and collaborators even produced any novel metaphors at all).

There are some caveats with the overall interpretation of the results of Experiments 1-4, specifically regarding the way in which Experiment 4 compares to Experiments 1 and 2: First, in Experiment 4 each participant saw only 2 critical items, whereas in Experiments 1 and 2 participants saw the full set of 36 items. It could therefore be the case that repeated exposure to the video primes interfered with an underlying true priming effect that our experimental set-up in Experiments 1 and 2 could not detect. It is also possible that the lack of an effect was due to the verbs being embedded in a sentence, regardless of whether the context encourages a literal or metaphoric interpretation of the verb. This is unlikely, considering that in Experiment 2 the Video-Prime and the verb were almost as temporally adjacent as in Experiment 4, but it cannot be ruled out completely. Further research is necessary in order to determine the exact nature of the prime-verb relation and the different contexts under which a priming effect of contained video on containment verbs could arise. Nevertheless, the experiments presented in this chapter can be seen as a step forward in understanding how metaphors are processed outside of the narrow realm of nominal metaphors.

Now I turn to the next critical issue: The question of symmetry of metaphoric processing. The following chapter presents two eye-tracking Experiments designed to address this issue using a modified version of the VWP.

*The fact that participants direct their visual attention towards objects in a concurrent scene whose names they hear is not particularly surprising. After all, one of the purposes of language is precisely to direct visual attention. More interesting (and perhaps more constraining, theoretically) are the conditions under which language-mediated eye movements occur towards objects that are not referred to in the language.*

— Altmann & Kamide, 2007, p. 511.

# 4

# Metaphoric (A)symmetry: the Case of Verb-Object Metaphors

In chapter two, I explained how one of the key differentiating predictions of categorization and Implicit Comparison views is whether or not the metaphoric topic and vehicle play the same role during the construction of figurative meaning.

The Implicit Comparison view sees both elements playing the same initial role: The structure of the elements needs to be aligned in order for inferences to be projected, meaning that initial processing has to be role-neutral. On the other hand, the Category Inclusion view stipulates role-specific tasks: The meaning of the metaphoric vehicle is contextually adjusted to create an ad hoc category, while the metaphoric topic provides the necessary parameters that steer the way in which the ad hoc category is constructed. In this chapter, I examine this question with empirical evidence from two experiments using the Visual World Paradigm.[1]

As mentioned in chapter two, a big problem with testing these predictions is the fact that the order of the elements of a nominal metaphor cannot be reversed without altering the meaning of the expression as a whole. Wolff and Gentner (2011) worked

[1]A version of this chapter presenting the results of Experiment 5 was published in Rodriguez Ronderos, Guerra, et al. (2020). Camilo Rodríguez Ronderos and Pia Knoeferle conceptualized the study. Camilo Rodríguez Ronderos created and normed the materials and conducted the experiment. Camilo Rodríguez Ronderos and Ernesto Guerra analyzed the data. Camilo Rodríguez Ronderos wrote the paper with feedback from Pia Knoeferle and Ernesto Guerra.

around this constraint by showing that participants in two experiments accepted reversed metaphors (*a virus is a rumor*) as meaningful at the earliest time intervals (500 and 600 milliseconds) in a way that was not significantly different from the way in which they accepted forward metaphors (*a rumor is a virus*) as a meaningful. This is consistent with an initial alignment stage, in which the structure of the elements of a metaphor is aligned regardless of their relative position. However, this piece of evidence does not tell us anything about the way in which participants incrementally construct a mental representation of the metaphoric expressions as a function of the order of the elements. As discussed in chapter two, it also does not take into consideration the way that the relationship between topic and vehicle is influenced by the preceding linguistic context. These two facts make it necessary to find better testing grounds for this particular hypothesis.

One way forward is to look for metaphoric constructions that do allow for a felicitous reversal of the order of their elements (outside the limited scope of English nominal metaphors), and that do consider the effect of the preceding context on comprehension. Such constructions occur in German in cases in which the metaphoric vehicle appears as the accusative object of a sentence, as in (1) below:

*(1a) Sebastian füttert* TOPIC *eine Prinzessin* VEHICLE.

"Sebastian feeds TOPIC a princess VEHICLE"

*(1b) Sebastian wird eine Prinzessin* VEHICLE *füttern* TOPIC.

"Sebastian will a princess VEHICLE feed TOPIC"

An important feature of German syntax is that it alternates from an SVO to an SOV surface sentence structure in the presence of an auxiliary verb. From an incremental semantic processing point of view, this means that our mental representations can be constructed differently, depending on which information comes first: the main verb (1a) or the accusative object (1b). This difference is ideal for testing the contribution that each element of the metaphor (the vehicle or the verbal topic) is making by reversing the order of presentation and examining the activation patterns of the figurative and literal meanings of the expression.

It's important to note that the metaphor in (1) does not have the canonical topic-vehicle structure present in metaphors such as my lawyer TOPIC is a shark VEHICLE. In (1), there is no explicit mention of the metaphoric topic. Instead, the verb *füttert* can be seen as the "verbal topic", given that it is the only element of this expression that provides information about the nominal topic of the metaphor: *Füttert*, embedded in an appropriate relevant context (see (2b below)), will most likely activate the metaphoric topic *Katze*, given that the verb provides enough semantic and contextual constraints for participants to anticipate its most likely referent.

The idea that *füttert* will activate the metaphoric topic *Katze* given a supporting visual context is supported by research on incremental language processing (see chapter two for a brief introduction). Specifically, research on the anticipation of post-verbal objects has shown that the information provided by a verb can be reliably used to generate expectations about the semantic features of an upcoming object. Altmann and Kamide (1999) found that when participants hear sentences that include verbs with restricting thematic roles (such as *eats* in the sentence *the boy eats cake*) participants' eye movements anticipated the direct object by moving towards the only depiction of an edible object in the given visual array upon verb onset presentation.

This effect has been replicated successfully with larger samples (Hintz et al., 2017) and 2-year-old children (Mani & Huettig, 2012). Importantly, Kamide et al. (2003) found that even when a verb does not provide enough constraining information on its own (such as *ride*), participants incorporate their knowledge about the sentential subject (*the girl will ride…* vs. *the man will ride…*) to anticipate an upcoming object based on an available visual context (more looks to a visually represented motorbike when *the man* was the subject compared to when the subject was *the girl*).

This line of research raises the question of whether anticipating an upcoming verbal object is driven by verb semantics alone or by the contextual plausibility granted by the previous linguistic context. Metusalem et al. (2012) asked a related question in an investigation of the role of event knowledge during incremental language processing. Their goal was to show that information about an event provided by a (linguistic) context could activate information that is compatible with the event at hand yet incompatible with the unfolding linguistic input. They found that words that are contextually incompatible can nevertheless elicit a reduced N400 effect if they can be said to belong to the general domain of the event at hand ("semantically related"), compared to contextually incompatible and semantically anomalous words. This finding would suggest that regardless of grammatical and thematic fit, information that is relevant to the overall construction of a mental representation of an event can facilitate processing. But what would happen if the information provided by the verb was at odds with the information provided by a linguistic context? Would participants' ability to anticipate an object be impaired or would they preferentially rely on verbal or contextual expectations during processing?

The experiments in this chapter are designed to answer this question as well as the question of the role-specificity of the elements of a metaphor: In (1) The verb is biased towards expecting the verbal object to be an animal (or perhaps a toddler), since the verb *füttern* carries this information as part of its thematic role constraints. But, (1) can be preceded by two different contexts, such as (2a) and (2b) below, that seem to shift the expectations for the upcoming verbal object from *Katze* (an animal), to *Prinzessin* (a human), regardless of the thematic role assignment constraints that might be associated with the verb *füttern* on its own:

(2a, literal context)

*Sebastian liebt eine berühmte Adlige. Er hat sie in einem Schloss kennengelernt und seitdem sind sie unzertrennlich. Die Adlige ist schwach und abhängig, und kann sehr hilfsbedürftig sein. Deswegen tut Sebastian alles für sie, wenn sie Hunger hat. Er wird sich immer um sie kümmern wollen.*

"Sebastian loves a famous noble woman. He met her in a castle and they have been inseparable since. The noble woman is weak and dependent and can be very needy. That's why Sebastian would do anything for her when she's hungry. He will always want to take care of her."

(2b, metaphoric context)

*Sebastian liebt eine wunderschöne Katze. Er hat sie in einem Tierheim adoptiert und seitdem sind sie unzertrennlich. Die Katze ist verwöhnt und launisch, und kann sehr wählerisch sein. Deswegen würde Sebastian alles für sie tun, wenn sie Hunger hat. Er wird sich immer um sie kümmern wollen.*

"Sebastian loves a beautiful cat. He adopted her in a shelter and they have since been inseparable. The cat is spoiled and moody and can be very fussy. That's why Sebastian would do anything for her when she's hungry. He will always want to take care of her."

Two things can be said about the change in interpretation brought on by the different preceding contexts:

(a) Verb-Object metaphors such as (1) are highly sensitive to the preceding context: Without the relevant contextual information of what the verb might be referring to, one should be less likely to understand that *Prinzessin* is in fact a spoiled cat. More generally, given that the meaning of novel metaphors is not retrieved from memory but constructed on the fly for every individual expression, the relationship between a metaphor and its context is crucial in order to determine if a metaphor is felicitous and if can be processed easily.

(b) Contextually derived expectations might be able to override the on-line expectations of thematic roles generated by a verb alone. This would suggest that the thematic role constraints of a verb might not play the most important role during sentence processing. Instead, the overall construction of a mental representation of an event might be what determines how post-verbal object anticipation operates, in line with the findings of Metusalem et al. (2012).

The overall goal of the two VWP experiments in this chapter is to examine how German verb-object constructions are processed as a function of the relative position of the verb and direct object. In Experiment 5, verb-object constructions were embedded in a linguistic context that either biased towards a novel metaphoric or a literal interpretation of the verb's direct object. Participants then saw a visual grid with four pictures and had to click on the picture that best represented the story after hearing the target utterance. In Experiment 6, the target utterance was not preceded by any linguistic context at all.

These two experiments will help us understand (i) what the elements of a metaphoric expression contribute to processing (and when they do so), (ii) how interpretation biases (literal vs. metaphoric) imposed by a preceding linguistic context interact with the verb's thematic role constraints, and (iii) what the role of context is when deriving metaphoric vs. literal interpretations of German verb-object constructions.

## 4.1 Experiment 5

### 4.1.1 Participants

Thirty-two native speakers of German (aged 18 to 31) with normal or corrected-to-normal vision gave their informed consent and received 12 Euros each for their participation. The number of participants was determined via an a-priori power analysis through simulations using data from a pilot study (see Appendix C for details of the pilot study) with the help of the R package SimR(Green & MacLeod, 2016). The power analysis determined that with 32 participants statistical power would be above 80% assuming a true effect size (for the difference between Early-Metaphoric and Late-Metaphoric in the vehicle region) with a Cohen's d value of 0.2. All participants were right handed and had normal or corrected-to-normal vision. This study was covered by the ethics vote of the psycholinguistics lab of the Humboldt-Universität zu Berlin.

### 4.1.2 Materials and design

#### Linguistic stimuli

For Experiments 5 and 6, 36 verb-object metaphors were created. In Experiment 5, these were paired with a literal and a metaphorical context (as seen in Table 4.1 ). The 36 items were selected from an original pool of 38 according to the results of a ratings task conducted with an independent group of participants (see norming study 1 below). The literal and metaphoric contexts and target utterances were matched for length within every item ($\pm 2$ characters) as well as for syntactic structure. The last sentence in each item was identical across conditions with the exception of the disambiguating word: In the literal conditions, participants heard the literal disambiguating word (*Adlige*, "noble woman") and in the metaphoric contexts they heard the metaphoric disambiguating word (*Katze*, "cat") (see Table 4.2). This word was always the same word used in the written context to describe the given referent. After the norming tasks were completed, all sentences were recorded by a female native speaker of German. All critical items are included in Appendix B.

**Table 4.1:** Example of a critical linguistc context in Experiment 5

| **Condition** | **Item** |
|---|---|
| Literal context | Sebastian liebt eine berühmte Adlige. Er hat sie in einem Schloss kennengelernt und seitdem sind sie unzertrennlich. Die Adlige ist schwach und abhängig, und kann sehr hilfsbedürftig sein. Deswegen tut Sebastian alles für sie, wenn sie Hunger hat. Er wird sich immer um sie kümmern wollen. |
| *English Translation* | *"Sebastian loves a famous noble woman. He met her in a castle and they have been inseparable since. The noble woman is weak and dependent and can be very needy. That's why Sebastian would do anything for her when she's hungry. He will always want to take care of her."* |
| Metaphoric Context | Sebastian liebt eine wunderschöne Katze. Er hat sie in einem Tierheim adoptiert und seitdem sind sie unzertrennlich. Die Katze ist verwöhnt und launisch, und kann sehr wählerisch sein. Deswegen würde Sebastian alles für sie tun, wenn sie Hunger hat. Er wird sich immer um sie kümmern wollen. |

| Condition | Item |
|---|---|
| *English translation* | *"Sebastian loves a beautiful cat. He adopted her in a shelter and they have since been inseparable. The cat is spoiled and moody and can be very fussy. That's why Sebastian would do anything for her when she's hungry. He will always want to take care of her."* |

**Table 4.2:** Example of a target utterance in Experiments 5 & 6

| Condition | Item |
|---|---|
| Early verb position, literal context | Sebastian füttert$_{\text{topic}}$ eine Prinzessin$_{\text{vehicle}}$, und wird unablässig der Adligen beistehen. |
| *English Translation* | *"Sebastian feeds a princess and will relentlessly the noble woman support"* |
| Early verb position, metaphoric context | Sebastian füttert$_{\text{topic}}$ eine Prinzessin$_{\text{vehicle}}$, und wird unablässig der Katze beistehen. |
| *English Translation* | *"Sebastian will a princess feed and will relentlessly the cat support"* |
| Late verb position, literal context | Sebastian wird eine Prinzessin$_{\text{vehicle}}$ füttern$_{\text{topic}}$, und wird unablässig der Adligen beistehen. |
| *English Translation* | *"Sebastian will a princess feed and will relentlessly the noble woman support"* |
| Late verb position, metaphoric context | Sebastian wird eine Prinzessin$_{\text{vehicle}}$ füttern$_{\text{topic}}$, und wird unablässig der Katze beistehen. |
| *English Translation* | *"Sebastian will a princess feed and will relentlessly the cat support"* |

### Visual stimuli

For every item, a grid of 4 pictures was created: One depicting the literal meaning of the metaphoric vehicle, one the metaphoric meaning and two unrelated distractor images. The images used were chosen from a pool of freely available photographs found on a popular search engine. Figure 4.1 shows the visual grid created for the target item in Table 4.2. The top-right image depicts the metaphoric interpretation of the vehicle (*Prinzessin* understood as a cat), while the bottom-left image depicts the literal interpretation (*Prinzessin* understood as a princess). The position of the images was randomized across items and participants.

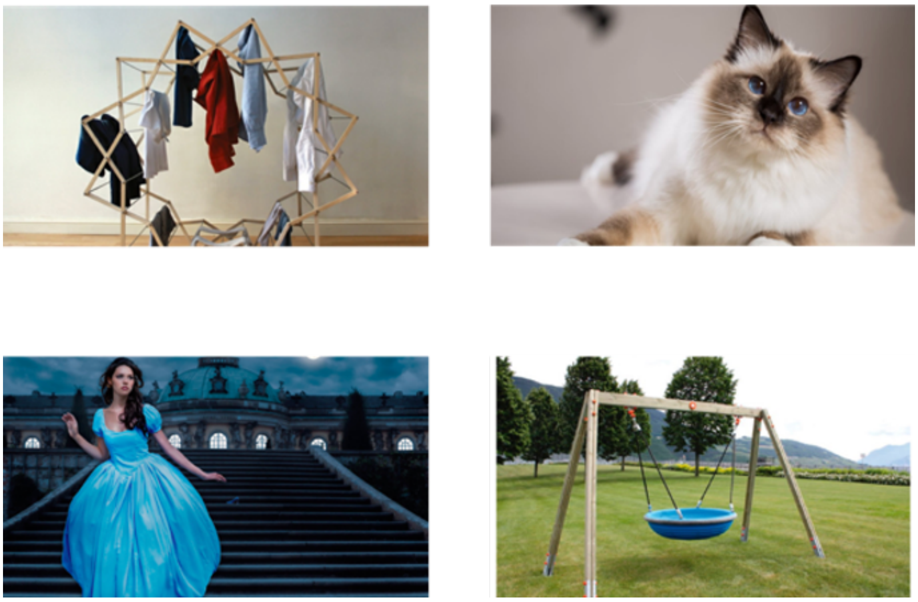

**Figure 4.1:** Example of a visual grid for a critical item in Experiments 5 & 6

### Norming study 1: metaphor aptness and familiarity

20 monolingual German native speakers (ages 18-31) rated the goodness of the metaphoric expressions given their context on a scale from 1 (meaning “incomprehensible”) to 7 (meaning “perfectly comprehensible”) after having to choose from 3 possible sentences the correct intended metaphoric meaning. This was meant to ensure that the created novel metaphors could actually be easily understood. Items with a mean aptness score below 4 were dropped from the main experiment.

Additionally, participants also rated the familiarity of the metaphors, by indicating whether they had ever heard the target metaphoric vehicle used in this specific metaphoric meaning before on a scale from 1 (meaning “never before”) to 7 (meaning “very often”).

The aptness results showed that every item was assigned its correct meaning over 80% of the times, with the exception of items 1 (52%) and item 24 (75%). These two items were also the only ones to be rated below 4 by participants (Item 1: 2.63; item 24:3.35). They were therefore excluded from the main Experiment. The remaining 36 items, which were used in Experiments 5 and 6 as well as 11-14 (presented in chapter five) are included in Appendix B.

### Norming study 2: Gated completion task

In order to assess whether the verbs chosen for the study could indeed generate expectations about possible upcoming objects independently from context, 20 monolingual native speakers of German (ages 18-31) were asked to take part in a gated sentence-picture completion task. None of the participants had previously taken part in the first norming task, nor did they take part in the subsequent study presented in this chapter.

Participants read only subject and verb of the 36 critical items on a computer screen. These incomplete sentences were visible for 2 seconds, after which they disappeared from the screen and a visual grid appeared containing the same four pictures that were to be shown in Experiment 5 (see Figure 4.1 ): Target, competitor, and two distractor images. The goal was to make sure that the subject-verb pairings could elicit a consistent expectation of the target picture (the cat, in Figure 4.1 ) as the upcoming object. Participants had a maximum of three seconds to select the picture that they thought would best describe a possible direct object of the sentence.

The results of this norming task were used to adjust both pictures and verbs of those items in which the target picture was selected less than 75% of the times. The pretest was then re-run with a different set of 15 monolingual German native speakers who did not take part in any other of the experiments in this chapter.

The results of this second (and final) norming study showed that the selected sentences together with the selected pictures generated strong expectations regarding upcoming verbal objects: For every critical item, participants selected the target image as the appropriate object over 75% of the times. Final results of the norming study are shown in Figure 4.2 .

### Filler items

72 combinations of written contexts + spoken utterances were created and used as filler items. These included metaphoric utterances, idiomatic, and literal sentences. For the filler trials, there was always one target image and three distractors, so that

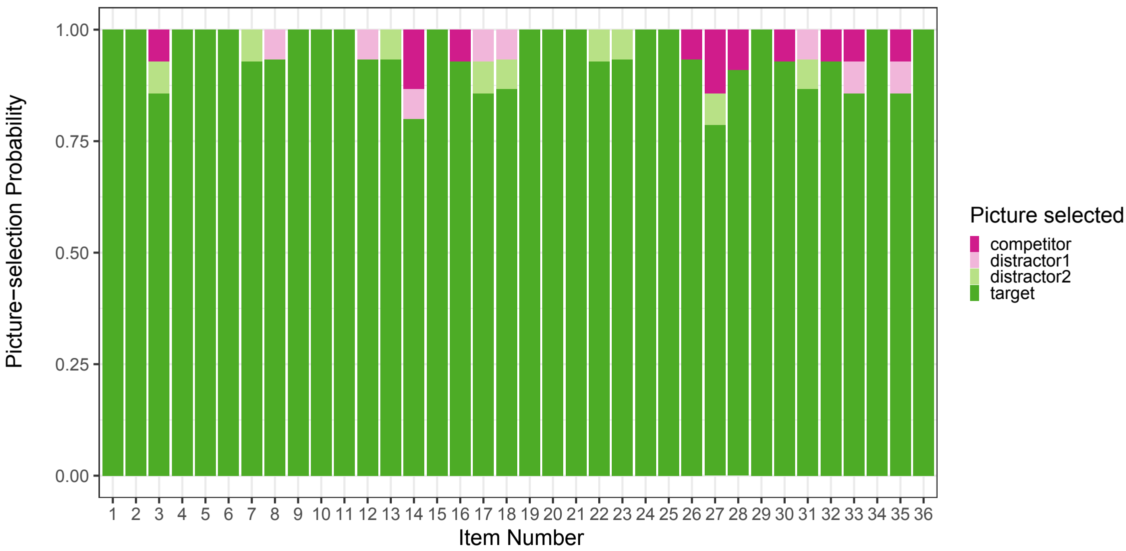


**Figure 4.2:** Summary of results of the picture-selection norming task of the materials of Experiment 5

participants could easily and reliably establish the appropriate referent. Item (3) below is an example of a filler item. All filler items are included in Appendix B.

**Design**

The experiment had a 2x2, repeated-measures design with the factors Contextual Bias (literal vs metaphoric) and Verb Position (early vs. late). "Contextual Bias" refers to the type of linguistic context that participants read prior to hearing the target utterance. "Verb Position" describes whether the target utterance heard had its lexical verb before or after the direct object. This resulted in four different versions of every item for each of the four experimental conditions, as seen in Tables Table 4.1 and 4.2 .

### 4.1.3 Predictions

The higher-order prediction of Experiment 5 refer directly to the theoretical debate presented in chapter two: If both elements (verbal-topic and vehicle) of the metaphor contribute equally to the construction of the metaphoric meaning (as per Wolff & Gentner, 2011), then we should expect the pattern of activation of literal and figurative meaning to be similar regardless of the order of presentation of the elements, i.e. regardless of whether the verb precedes or follows its direct object. If, on the other hand, each element has a role-specific contribution (as per Glucksberg, 2001, p. 55), we should see different activation patterns depending on the order of presentation.

Concretely, if the Implicit Comparison view holds, we should find that when participants hear the metaphoric vehicle (*Prinzessin*), the relationship between looks to the princess (literal picture) and the cat (metaphoric picture) should be similar in the early verb and late verb constraint conditions. If, on the other hand, the Category Inclusion view holds, we should find differences in this relationship across conditions. In particular, we should find that when participants hear the topic prior to the vehicle, they should be able to settle on a metaphoric interpretation of the vehicle more easily upon hearing the metaphoric vehicle than when they hear the vehicle prior to the topic.

The second prediction for Experiment 5 concerns the type of information necessary for the anticipation of a verb's likely upcoming object. Altmann and Kamide (1999) showed that when hearing a verb, participants preferentially directed their gaze to a picture of an object that is most typically associated with this verb. But what happens when the information provided by a linguistic context suggests that the upcoming verbal object will be a different one to that most typically associated with the verb?

This is the case of the literal early-verb condition: the linguistic context guides participants to expect the visual representation of the literal princess to be the upcoming referent, but the verb (*füttert*) suggests that the image of the cat should be the most likely referent, as can be seen by the results of the second norming study. There are (at least) three possible scenarios: Participants could rely on (i) the verb or (ii) on the context to anticipate an upcoming object, or they could (iii) not anticipate any object whatsoever. Accordingly, in the literal early-verb condition, we would either expect (i) a preference for looking at the cat (compared to the princess) or (ii) a preference for looking at the princess (compared to the cat), or (iii) no significant preference for either one in the verb region.

### 4.1.4 Procedure

Participants' eye movements were recorded using an Eyelink 1000 plus desktop head-stabilized tracker, produced by SR Research. At the beginning of each experimental session, the eye-tracker was calibrated with a 9-point calibration procedure to ensure accurate monitoring of the right eye, which was the only one tracked. The procedure was performed and repeated until there was less than a maximum error of 1 degree. If it was not possible to meet this criterion, the experiment was aborted and participants were replaced. Re-calibration was performed after every pause in the Experiment, i.e. twice more. After calibration, participants saw three practice trials before the experiment began.

Each trial in the experiment consisted of five phases (see Figure 4.3): First, participants read a 4-sentence context (see Table 4.1 ) and clicked anywhere on the screen when they were ready to continue. They then saw 4 pictures on the screen: One representing the literal (princess) and another the metaphoric (cat) meaning of the target sentence as well as two distractor images (see Figure 4.1 ). After two seconds of image preview, participants heard the target utterance while the pictures remained on the screen (see Table 4.2 for an example of a critical item). Their task was to click on the image that they thought best fit both the written context and the spoken sentence. They could only move the mouse once the utterance had been played in its entirety. On 1/3 of the trials (36 trials, filler trials only) they answered multiple choice questions on the content of either the written text (12 trials), the pictures (12 trials), or the spoken utterance (12 trials).

### 4.1.5 Analysis

I quantified participants' viewing preference by measuring the log-ratio (Arai et al., 2007) between looks to the metaphoric picture divided by looks to the literal picture when participants heard the topic (*füttert* or *füttern*, "feeds") and the vehicle (*Prinzessin*, "princess"). Log-ratio values are centered around 0, where

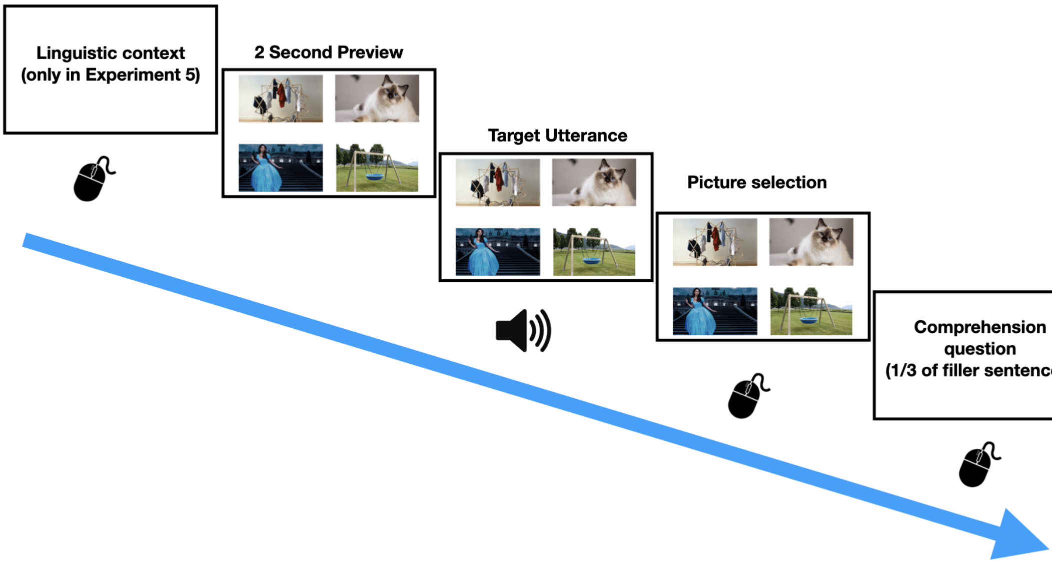


**Figure 4.3:** Example of the progresssion of a trial in experiments 5-6

positive values indicate a preference for the metaphoric picture and negative values a preference for the literal picture.

Additionally, I measured log-ratio values for the parts of the sentence that followed the metaphoric expression by dividing it into three different regions of interest. This resulted in the fitting of 5 different statistical models on 5 different sentence regions. The regions can be seen in Table 4.3.

**Table 4.3:** Regions of interest for a critical item in Experiments 5 & 6

| Region | Example Item | Comment |
|---|---|---|
| **Vehicle** | *eine Prinzessin* ("A Princess") | The position of this region varied between conditions |
| **Verb** | *füttert* (verb-early conditions), or *füttern* (verb-late conditions) ("feeds") | position in sentence varied between conditions |
| **UND** | *und wird* ("and will") | This region was identical across all critical items, position in sentence identical within every item |
| **ADV** | *unablässig* ("relentlessly") | This region was identical within every item (it was always an adverb), position in sentence was identical within every item |

| Region | Example Item | Comment |
|---|---|---|
| **DIS** | *der Katze* (metaphoric conditions)(“the cat”) or *der Adligen* (literal conditions) (“the noble woman”) | Word is identical to the word in the written context, position in sentence identical within every item, word changes according to experimental condition. Word disambiguates the intended referent (thus the name, DIS). |

This specific distribution of regions for analysis is different to the one planned before the Experiment was run. The original planned analysis compared the first and second element in each condition to the element in the other condition that had the same position: The first element in the verb-early conditions (*füttert*) was compared to the first element in the verb-late conditions (*Prinzessin*) and the second element in the verb-early conditions (*Prinzessin*) was compared to the second element in the verb-late conditions (*füttern*). I chose to deviate from the original analysis because I considered that the one presented here was a more appropriate direct comparison seeing that it compares equal to equal (i.e. comparing hearing the word *Prinzessin* in the verb-early condition to hearing the word *Prinzessin* in the verb-late condition). This allows for a more straightforward interpretation of the results as well as for the fitting of more appropriate statistical models. The change in analysis does not change the conceptual predictions for the experiment nor the interpretation of the results. The original analysis and results are presented in detail in Appendix C.

I fitted five linear mixed-effects models to the data, one for each one of the five regions of interest illustrated in Table 4.3. The model-fitting process followed the recommendations of Barr et al. (2013). All models included the factors Contextual Bias, Verb Position and their interaction as fixed effects as well as trial number as a control variable.

Models were coded using a treatment contrast scheme in order to tailor the contrasts to specific hypotheses of the experiment, as recommended by Schad et al. (2020). When using treatment contrast, one condition is used as a baseline (the intercept of the model), with the coefficients of the predictor variables representing direct comparisons between a predictor and the baseline group while keeping all other predictors at a fixed level. This differs from the ANOVA-style, sum-contrast traditionally used (which was used to analyze the experiments in the previous chapter), in which the intercept represents the overall mean of all conditions, and the coefficients of the predictor variables represent the effect of a specific factor compared to the overall mean. Further details on why this specific contrast-coding scheme was used can be found in Appendix C.

Importantly, when using treatment contrast the coefficient of the intercept tests the null-hypothesis of whether the outcome value (i.e., log-ratio) is equal to zero. This means that the intercept term will tell us if there is a preference for either literal (negative log-ratio) or metaphoric (positive log-ratio) picture in the condition chosen as the baseline judging by whether or not the intercept is significantly different from zero.

As in the previous chapter, all tables reporting the results of regression analyses are structured as follows: For every variable, the tables report the regression coefficient, followed by the confidence intervals (in brackets) and the t-value. Stars on the t-values show the statistical significance cut-offs, based on the variable's p-value.

### 4.1.6 Results

#### Non-gaze Data

**Context reading times** To check if there were any differences in reading times of the contexts prior to listening the target sentence, a linear mixed-effects model was fitted to the data using the square root of the reading times as the dependent measure. This dependent measure was chosen following the results of a Box-Cox test. The model showed no significant differences between reading times of the two types of contexts. The results are shown in Figure 4.4.

**Picture Selection Times** For all trials where participants selected the correct picture, I analyzed the time it took them to select said picture starting from the offset of the target utterance. A linear mixed-effects model was fitted to the picture selection times. The picture-selection times were transformed using an inverse square root transformation, following the results of a Box-Cox test. The statistical model fitted on the data included trial order as a control variable and random intercepts and slopes for both main effects (context bias and verb position) and their interaction by items and subjects. This model showed no significant differences between conditions. The results are shown in Figure 4.5.

**Comprehension questions** The goal of including comprehension questions on 1/3 of all filler items (i.e. in 24 trials of the experiment) was to make sure that participants remained attentive throughout the entire experimental session. The results show that participants were overall very accurate at answering comprehension questions. The mean response accuracy was 93%, with median accuracy being 94%. The minimum score for an individual participant was 84% and the maximum was 100%. Based on these scores, no participants were excluded from the analysis.

**Picture Selection** Of the total 1150 trials in the data, participants clicked on the correct target picture 1086 times. The 64 times they clicked on a picture other than the target were all found in the metaphoric conditions. This means that overall accuracy was above 99%, and accuracy in the metaphoric conditions above 98%.

**Gaze Data**

**Vehicle Region** The model for the vehicle region included random slopes for context bias by subjects and context bias and verb constraint by items. This model was fitted three times: the first one coding the metaphoric-early condition as the intercept, the second one doing the same with the metaphoric-late condition and the third one with the literal-early condition. This was done in order to asses the significance values of different conditions when these were coded as the intercept in the model. All models can be seen in detail in Table 4.4 .

The first version of the model (metaphoric-early condition as intercept) showed a significant difference between early and late metaphoric conditions. Additionally, the intercept of the model was positive and significantly different from zero. This

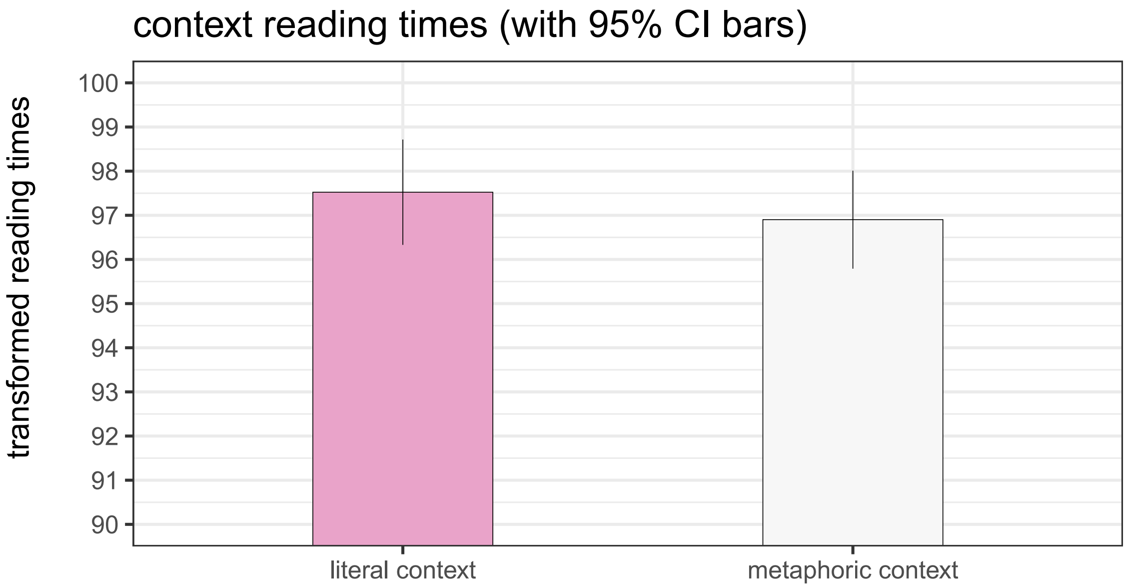


**Figure 4.4:** Summary of reading times of linguistic context, Experiment 5.

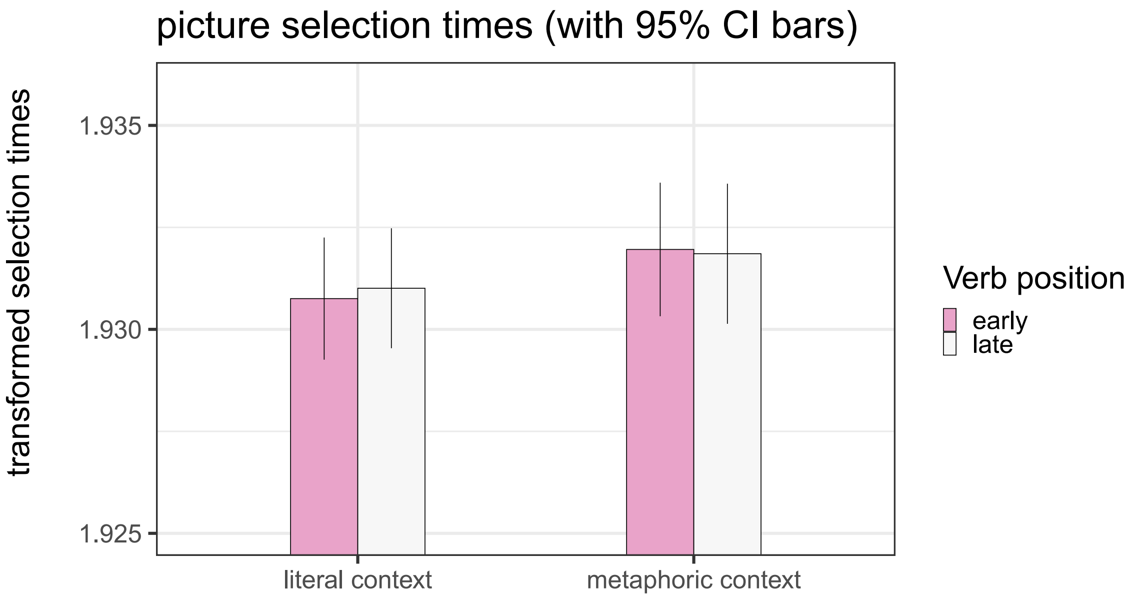


**Figure 4.5:** Summary of picture-selection times, Experiment 5.

result suggests that there is a difference in viewing pattern when hearing the vehicle depending on whether the vehicle appeared after or before the verb. The second version of the model (metaphoric-late as intercept) showed a significant difference between metaphoric-late and literal-late. The intercept of this model was negative and significantly different from zero. This suggests that when they heard the vehicle prior to the verb, participants mostly considered the literal picture (the princess) as the likely referent, whereas they mostly considered the metaphoric picture when hearing the vehicle after the verb. As for the literal conditions, the third model showed no significant difference between literal-early and literal-late conditions. This finding suggests that the gaze preference differences between early and late conditions found in this region were specific to metaphor comprehension.

**Verb Region** The model for the verb region included random intercepts and random slopes for both factors by items and random slopes for both factors by subjects. This model was fitted three times, analogously to the model of the vehicle region. The models are shown in Table 4.5 .

The first version of the model (metaphoric-early condition as intercept) showed a significant difference between metaphoric-early and literal-early and a significant interaction between both factors. The intercept of the model was positive and significantly different from zero, signifying an overall preference for the metaphoric picture throughout the region in the metaphoric-early condition. This suggests that when hearing the verb prior to the vehicle after having read a metaphorically biasing context, participants anticipated the object that was compatible with both context and the verb's selectional preferences (the cat).

The second version of the model (metaphoric-late as intercept) showed a significant difference between metaphoric-late and literal-late conditions. Furthermore, the intercept of the model was significantly positive, suggesting that participants quickly re-evaluated their assessment from the previous region and preferred the "cat" as the most likely referent.

Finally, the third model (literal-early condition as intercept) showed a significant difference between early and late literal conditions. The intercept of the model was not significantly different from zero, suggesting that participants did not reliably anticipate the object that was compatible with context (the princess) or the object compatible with the verb's selectional preferences (the cat) and only preferred to look at the princess when they heard the verb after the object (Late-Literal conditions).

**UND Region** The model for the "und" region included random intercepts and random slopes for both factors by items and random slopes for both factors by subjects. As with the previous regions, the model was fitted three times. Overall, the results of this region show that participants settled for a literal interpretation in both literal conditions and on a metaphoric interpretation in the metaphoric conditions: There were significant differences between both metaphoric conditions and their literal counterparts. All model intercepts were significant. The results are shown in Table 4.6 and can be visualized in Figure 4.8 .

**ADV Region** The model for the "adv" region included random intercepts and random slopes for both factors by items and random slopes for both factors by subjects. As with the previous regions, the model was fitted three times. As can be seen in Figure 4.9, the results are very similar to those found in the previous region, with significant differences between metaphoric and literal conditions and no differences between early and late verb conditions. All model intercepts were significant. The results are shown in Table 4.7 .

**DIS region** The model for the "dis" region included random intercepts and random slopes for both factors by items and random slopes for both factors by subjects. The results for this region were similar to those found in the previous two regions, with more pronounced effects (see Figure 4.10 ). This suggests that participants looked at the appropriate referent when hearing the disambiguating word. The models are shown in Table 4.8 .

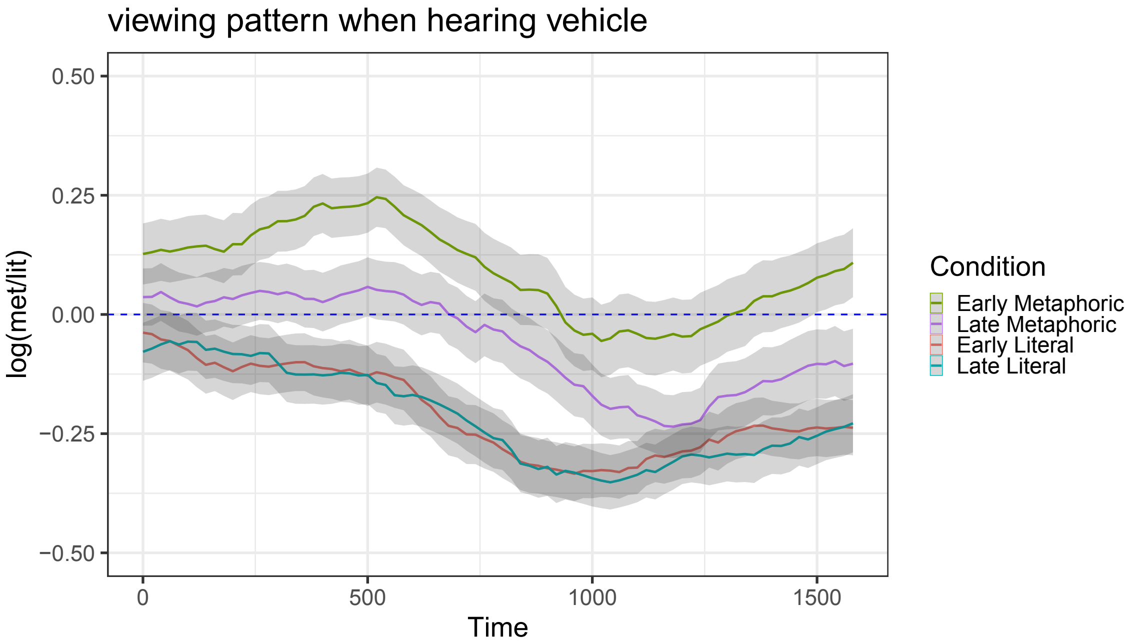


**Figure 4.6:** Summary of results for the vehicle region, Experiment 5

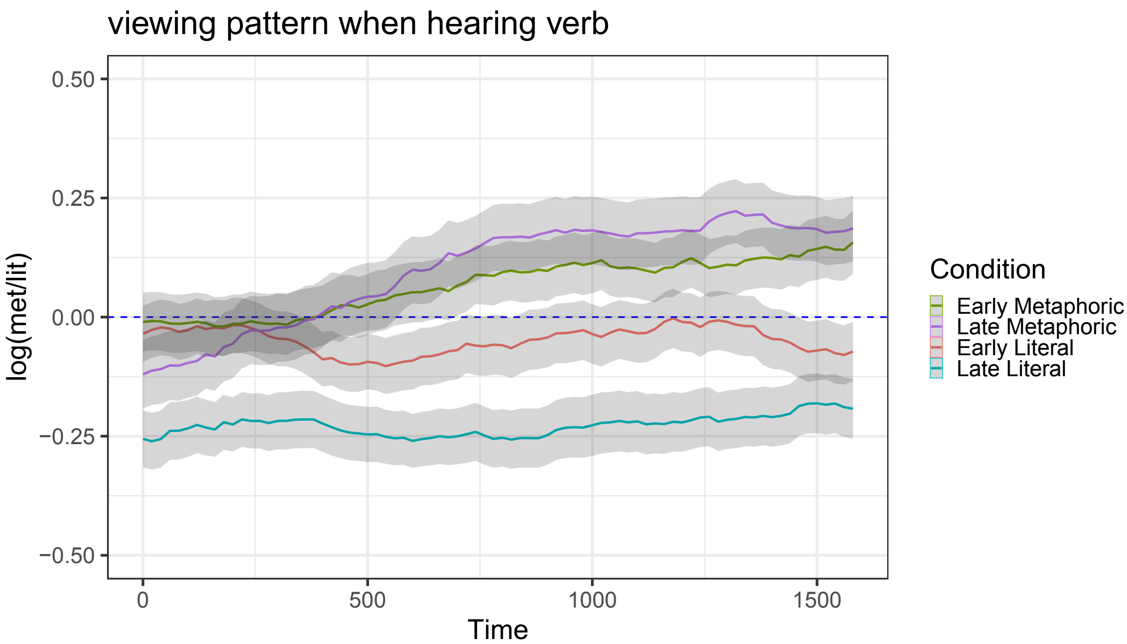


**Figure 4.7:** Summary of results for the verb region, Experiment 5.

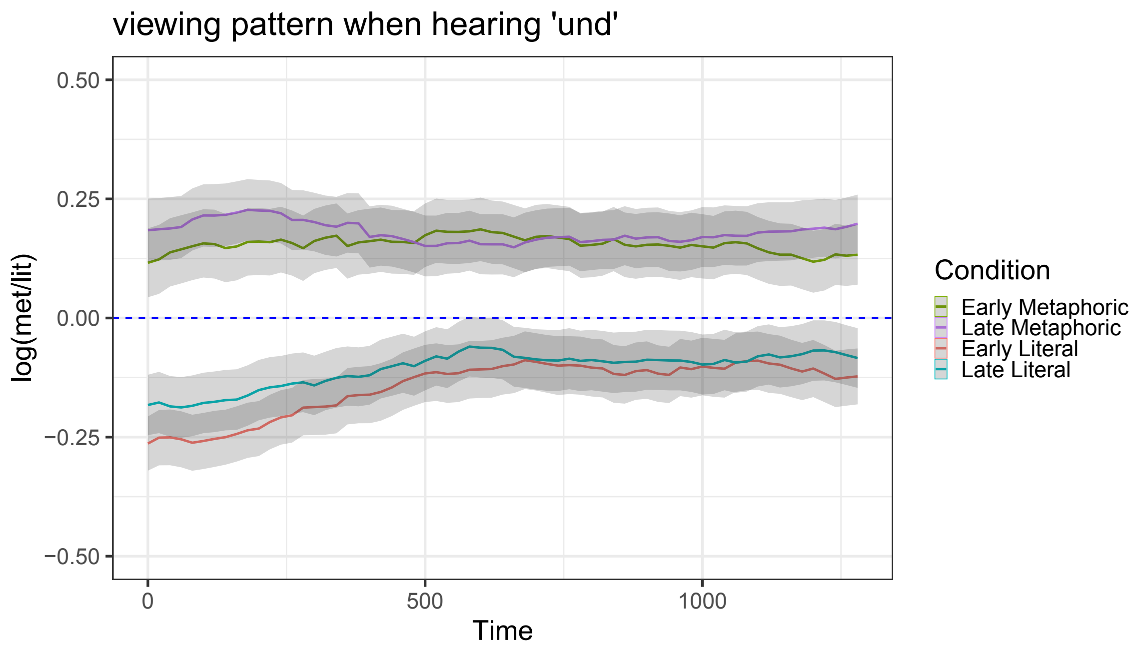


**Figure 4.8:** Summary of results for the und region, Experiment 5

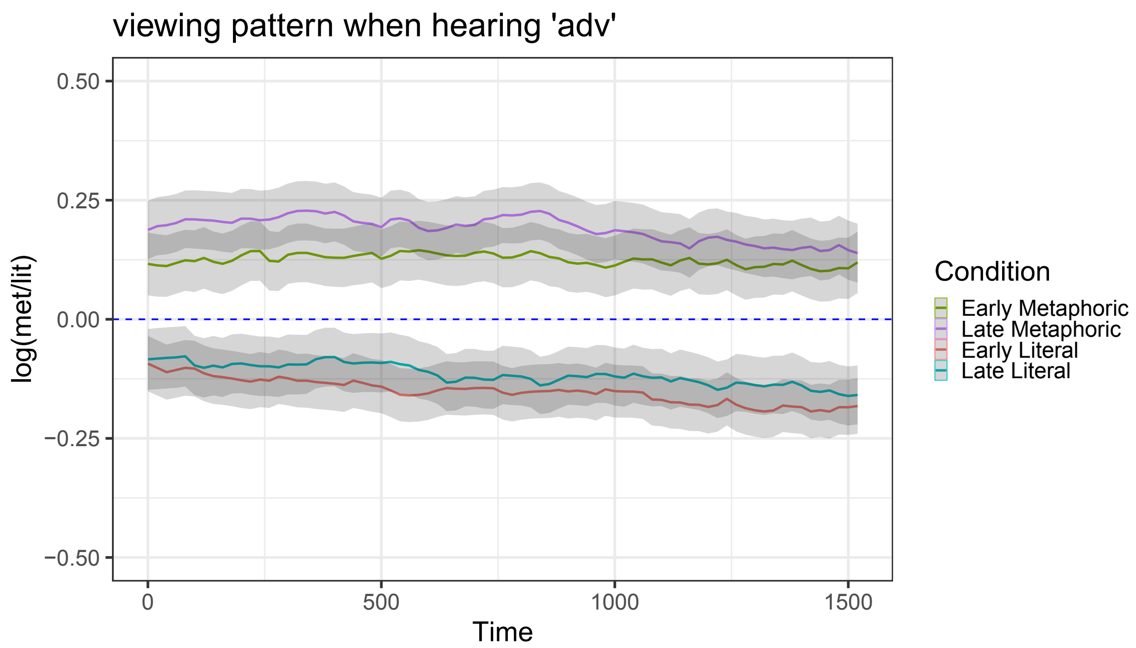


**Figure 4.9:** Summary of results for the adv, Experiment 5

**Table 4.4:** Regression analysis of log-gaze probability ratio in the vehicle region of Experiment 5

| | *Dependent variable:* logratio | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Early-Metaphoric vs. Early-Literal | $-0.304$<br>$(-0.380, -0.228)$<br>$t = -7.808^{***}$ | | |
| Early-Metaphoric vs. Late-Metaphoric | $-0.155$<br>$(-0.204, -0.107)$<br>$t = -6.288^{***}$ | | |
| Late-Metaphoric vs. Late-Literal | | $-0.154$<br>$(-0.230, -0.078)$<br>$t = -3.950^{***}$ | |
| Early-Literal vs. Late-Literal | | | $-0.005$<br>$(-0.054, 0.043)$<br>$t = -0.212$ |
| Trial Order | 0.001<br>(0.0002, 0.001)<br>$t = 3.125^{**}$ | 0.001<br>(0.0002, 0.001)<br>$t = 3.125^{**}$ | 0.001<br>(0.0002, 0.001)<br>$t = 3.125^{**}$ |
| Verb*Context Interaction | 0.150<br>(0.136, 0.164)<br>$t = 20.673^{***}$ | | |
| Intercept | 0.084<br>(0.039, 0.130)<br>$t = 3.619^{***}$ | $-0.071$<br>$(-0.117, -0.025)$<br>$t = -3.039^{**}$ | $-0.220$<br>$(-0.265, -0.174)$<br>$t = -9.427^{***}$ |
| condition at intercept | early-met | late-met | early-lit |
| Observations | 84,960 | 84,960 | 84,960 |
| Log Likelihood | −66,109.130 | −66,109.130 | −66,109.130 |
| Akaike Inf. Crit. | 132,236.300 | 132,236.300 | 132,236.300 |
| Bayesian Inf. Crit. | 132,320.400 | 132,320.400 | 132,320.400 |

*Note:* $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table 4.5:** Regression analysis of log-gaze probability ratio in the verb region of Experiment 5

| | *Dependent variable:* | | |
|---|---|---|---|
| | logratio | | |
| | (1) | (2) | (3) |
| Early-Metaphoric vs. Early-Literal | −0.122<br>(−0.204, −0.041)<br>t = −2.959** | | |
| Early-Metaphoric vs. Late-Metaphoric | 0.018<br>(−0.050, 0.085)<br>t = 0.509 | | |
| Late-Metaphoric vs. Late-Literal | | −0.311<br>(−0.392, −0.230)<br>t = −7.511*** | |
| Early-Literal vs. Late-Literal | | | −0.171<br>(−0.239, −0.103)<br>t = −4.947*** |
| Trial Order | −0.0001<br>(−0.0005, 0.0002)<br>t = −0.701 | −0.0001<br>(−0.0005, 0.0002)<br>t = −0.701 | −0.0001<br>(−0.0005, 0.0002)<br>t = −0.701 |
| Verb*Context Interaction | −0.189<br>(−0.202, −0.175)<br>t = −26.723*** | | |
| Intercept | 0.075<br>(0.015, 0.135)<br>t = 2.456* | 0.093<br>(0.033, 0.153)<br>t = 3.030** | −0.047<br>(−0.107, 0.013)<br>t = −1.547 |
| condition at intercept | early-met | late-met | early-lit |
| Observations | 85,200 | 85,200 | 85,200 |
| Log Likelihood | −63,613.150 | −63,613.150 | −63,613.150 |
| Akaike Inf. Crit. | 127,248.300 | 127,248.300 | 127,248.300 |
| Bayesian Inf. Crit. | 127,351.200 | 127,351.200 | 127,351.200 |

*Note:* *p<0.05; **p<0.01; ***p<0.001
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table 4.6:** Regression analysis of log-gaze probability ratio in the 'und' region of Experiment 5

| | *Dependent variable:* | | |
|---|---|---|---|
| | logratio | | |
| | (1) | (2) | (3) |
| Early-Metaphoric vs. Early-Literal | −0.303<br>(−0.360, −0.246)<br>t = −10.388*** | | |
| Early-Metaphoric vs. Late-Metaphoric | 0.026<br>(−0.013, 0.064)<br>t = 1.310 | | |
| Late-Metaphoric vs. Late-Literal | | −0.289<br>(−0.346, −0.231)<br>t = −9.886*** | |
| Early-Literal vs. Late-Literal | | | 0.040<br>(0.002, 0.078)<br>t = 2.057* |
| Trial Order | −0.001<br>(−0.001, −0.0005)<br>t = −4.305*** | −0.001<br>(−0.001, −0.0005)<br>t = −4.305*** | −0.001<br>(−0.001, −0.0005)<br>t = −4.305*** |
| Verb*Context Interaction | 0.015<br>(−0.001, 0.030)<br>t = 1.850 | | |
| Intercept | 0.174<br>(0.130, 0.219)<br>t = 7.752*** | 0.200<br>(0.156, 0.244)<br>t = 8.890*** | −0.129<br>(−0.173, −0.085)<br>t = −5.727*** |
| condition at intercept | early-met | late-met | early-lit |
| Observations | 70,590 | 70,590 | 70,590 |
| Log Likelihood | −54,084.480 | −54,084.480 | −54,084.480 |
| Akaike Inf. Crit. | 108,187.000 | 108,187.000 | 108,187.000 |
| Bayesian Inf. Crit. | 108,269.400 | 108,269.400 | 108,269.400 |

*Note:* *p<0.05; **p<0.01; ***p<0.001
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table 4.7:** Regression analysis of log-gaze probability ratio in the 'adv' region of Experiment 5

| | *Dependent variable:* | | |
|---|---|---|---|
| | logratio | | |
| | (1) | (2) | (3) |
| Early-Metaphoric vs. Early-Literal | −0.275<br>(−0.371, −0.180)<br>t = −5.668*** | | |
| Early-Metaphoric vs. Late-Metaphoric | 0.067<br>(0.003, 0.130)<br>t = 2.062* | | |
| Late-Metaphoric vs. Late-Literal | | −0.311<br>(−0.407, −0.216)<br>t = −6.407*** | |
| Early-Literal vs. Late-Literal | | | 0.031<br>(−0.033, 0.095)<br>t = 0.954 |
| Trial Order | −0.0004<br>(−0.001, −0.0001)<br>t = −2.474* | −0.0004<br>(−0.001, −0.0001)<br>t = −2.474* | −0.0004<br>(−0.001, −0.0001)<br>t = −2.474* |
| Verb*Context Interaction | −0.036<br>(−0.049, −0.023)<br>t = −5.346*** | | |
| Intercept | 0.132<br>(0.075, 0.190)<br>t = 4.498*** | 0.199<br>(0.142, 0.257)<br>t = 6.772*** | −0.143<br>(−0.201, −0.086)<br>t = −4.872*** |
| condition at intercept | early-met | late-met | early-lit |
| Observations | 83,622 | 83,622 | 83,622 |
| Log Likelihood | −57,894.060 | −57,894.060 | −57,894.060 |
| Akaike Inf. Crit. | 115,808.100 | 115,808.100 | 115,808.100 |
| Bayesian Inf. Crit. | 115,901.500 | 115,901.500 | 115,901.500 |

*Note:* *p<0.05; **p<0.01; ***p<0.001
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table 4.8:** Regression analysis of log-gaze probability ratio in the 'dis' region of Experiment 5

| | *Dependent variable:* | | |
|---|---|---|---|
| | logratio | | |
| | (1) | (2) | (3) |
| Early-Metaphoric vs. Early-Literal | −0.530<br>(−0.611, −0.448)<br>t = −12.775*** | | |
| Early-Metaphoric vs. Late-Metaphoric | 0.011<br>(−0.020, 0.042)<br>t = 0.702 | | |
| Late-Metaphoric vs. Late-Literal | | −0.553<br>(−0.634, −0.472)<br>t = −13.342*** | |
| Early-Literal vs. Late-Literal | | | −0.011<br>(−0.042, 0.020)<br>t = −0.706 |
| Trial Order | 0.001<br>(0.0005, 0.001)<br>t = 5.483*** | 0.001<br>(0.0005, 0.001)<br>t = 5.483*** | 0.001<br>(0.0005, 0.001)<br>t = 5.599*** |
| Verb*Context Interaction | −0.024<br>(−0.034, −0.013)<br>t = −4.535*** | | |
| Intercept | 0.245<br>(0.201, 0.289)<br>t = 10.984*** | 0.256<br>(0.212, 0.300)<br>t = 11.479*** | −0.285<br>(−0.333, −0.236)<br>t = −11.407*** |
| condition at intercept | early-met | late-met | early-lit |
| Observations | 134,728 | 134,728 | 134,728 |
| Log Likelihood | −91,040.380 | −91,040.380 | −90,017.050 |
| Akaike Inf. Crit. | 182,098.800 | 182,098.800 | 180,054.100 |
| Bayesian Inf. Crit. | 182,187.100 | 182,187.100 | 180,152.200 |
| *Note:* | *p<0.05; **p<0.01; ***p<0.001<br>values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) | | |

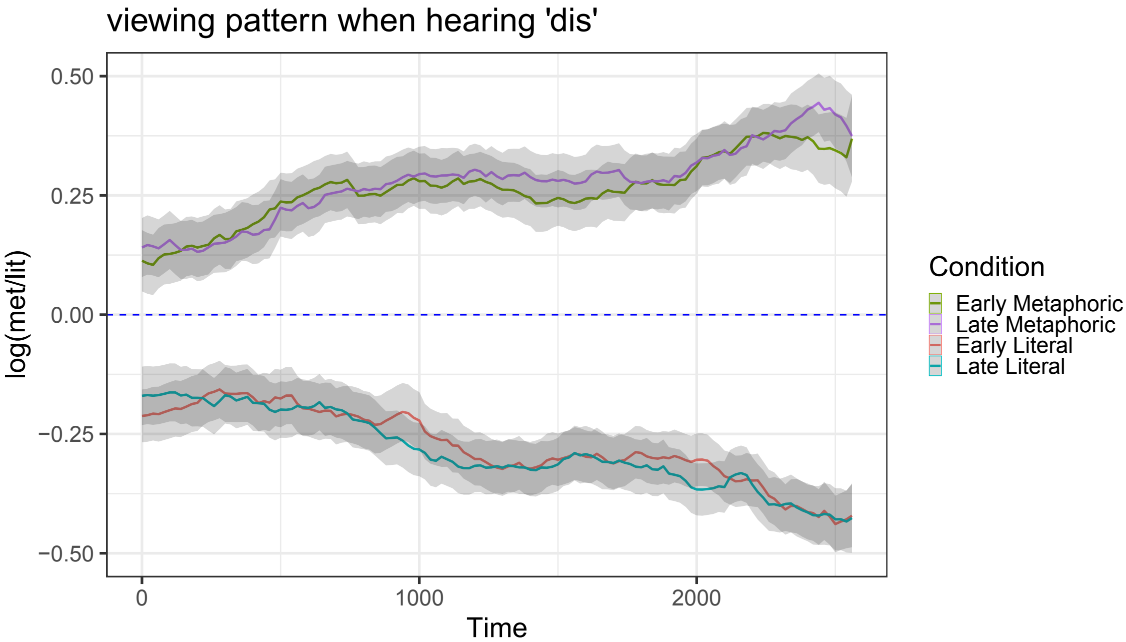


**Figure 4.10:** Summary of results for the dis, Experiment 5

### 4.1.7 Discussion

Experiment 5 investigated the time course of comprehension of German verb-object metaphors such as *Sebastian füttert eine Prinzessin* ("Sebastian feeds a princess"). *Prinzessin* could either be interpreted as a noble woman, or as a spoiled cat. This type of metaphor allows for the position of its elements to be reversed without altering the overall interpretation beyond a change in sentence tense ( *Sebastian wird eine Prinzessin füttern*). Moving the verb to the end of the clause means that as people hear *Prinzessin*, the topic (the verb) is not yet available. By contrast, for verb-second sentences, the topic verb has been encountered when participants process the vehicle (*Prinzessin*).

A written discourse context preceding the target utterance established a bias towards either a literal or a metaphoric interpretation of *Prinzessin*. In the target utterance, the verb established a bias towards the metaphoric interpretation (*füttert* prefers to take an animal as its object over an adult human being). This design permitted a juxtaposition of expectations caused by the linguistic context with the expectations associated with a verb's thematic role constraints (compatible vs. incompatible with the discourse context).

With regard to assessing the influence of individual verbs that were compatible (versus incompatible) with a preceding discourse context, Experiment 5 examined the anticipation of post-verbal objects when sentences are preceded by a linguistic context. Transitive verbs have been shown to generate expectations regarding their likely arguments in the absence of a linguistic context (Altmann & Kamide, 1999) and these expectations can direct visual attention to appropriate objects. However, it has also been shown that information that is semantically compatible with context yet incompatible with the unfolding linguistic input is activated during processing (Metusalem et al., 2012). In Experiment 5, participants read a context that biased towards expecting a specific referent (a princess or a cat) and then heard a sentence in which a verb (*füttert*, "feeds") constrained the upcoming referent via its thematic role constraints to only one of these referents (the cat).

The results, which can be visualized in Figure 4.7, show that in the case in which context expectations matched a verb's thematic role constraints (Early-Metaphoric condition, where both context and verb point towards the cat as likely referent), participants were able to anticipate a likely post-verbal referent when hearing the verb: They showed an overall viewing preference for the picture of the cat judging by the significantly positive log-ratio. When these sources of information (context and the verb's selectional preferences) were not aligned and participants heard the verb prior to the object (Early-Literal condition, where context points to the princess and the verb points to the cat as likely referent), no evidence for anticipation was found, given that the log-ratio in this condition was not significantly different from zero. This pattern suggests that in this case participants deployed a wait-and-see comprehension strategy instead.

The main finding of Experiment 5 concerns the asymmetry of processing verb-object metaphors. The current results show differences in the gaze patterns as a function of the position of the metaphoric elements and as a function of the contextual bias. Of particular interest here are the results of the Vehicle region, seen in Figure 4.6 . In the metaphoric conditions, the ratio of looks to the metaphoric picture (a cat) over looks to the literal picture (a princess) was larger if they had previously heard the verbal topic (*füttert*) compared to when they had not heard it yet. In fact, in the metaphoric early verb condition, there was an overall preference for fixating the metaphoric picture in the entire region, judging by the positive and significantly different from zero log-ratio in the metaphoric early verb condition, which can be seen in 4.4.

This finding is more in line with the Category Inclusion view, which assigns different roles to each element in a metaphor and predicts that the position of the elements will affect processing. Specifically, this view predicts that hearing a topic before the vehicle can produce a more rapid construction of metaphoric meaning upon hearing the vehicle, given that the topic provides a necessary set of parameters that allow for lexical modulation of the vehicle (See Chapter two for details).

The results are harder to account for by the Implicit Comparison view, according to which processing a metaphor requires structural alignment of the metaphor's elements prior to a projection of inferences. This alignment process is role-neutral and the same amount of information should be retrieved from the encoded meaning of the vehicle whether it appears prior to or after the topic. To accommodate the present results, it should be possible for the processing of the vehicle's encoded meaning to vary as a function of the position of the topic.

At this point, it is important to think about why participants would prefer to look at the depicted cat instead of the depicted princess when hearing the word *Prinzessin* ("princess") for the duration of the Vehicle region in the early-verb metaphoric condition. The strongest and most often found effect in the VWP literature is the referential effect: when hearing a noun, participants spontaneously and consistently look at the image that best represents the word they have just heard/are currently hearing (e.g. Cooper, 1974). So why would participants look at a cat when hearing *princess* if there was a literal princess available in the visual display?

For starters, it has been shown that referential effects are modulated by contextual fit: When embedded in an appropriate (sentential or discourse) context, participants can rapidly adjust the type of visual referent that best suits the unfolding language-mediated representation: Chambers et al. (2004) gave participants instructions requiring them to incorporate world knowledge information about the heard referents (hearing *pour the egg on the flour in the bowl* requires participants to know that the egg must be in liquid form). When hearing such sentences while seeing a visual grid in which there where two types of liquid eggs (one in a bowl and one in its shell), participants did not prefer one over the other when hearing the word *egg*. However, when one of these images was that of a solid egg, participants rapidly settled on the image of the liquid egg when hearing the word *egg*. This finding shows that the type of task at hand informs participants' viewing behavior and shapes their interpretation of the word *egg*, which results in a restriction of the possibility space of visual reference.

But a liquid egg is still a type of egg. A princess is not a cat. To explain the gaze pattern in the vehicle region of Experiment 5, it is necessary to think about the linking hypothesis of the VWP when used to study language processing. Altmann and Kamide (2007) (who provide the opening quote of this chapter) elaborated on this point in order to establish a more general way of accounting for a broad type of language-vision interactions. In two Experiments, Altmann and Kamide (2007) provided evidence showing that participants looked more at an empty wine glass after hearing *has drunk* compared to a full glass of beer. Conversely, they looked more at the full beer glass when hearing *will drink* compared to the empty wine glass. This finding extends the authors' previous findings by showing that verb morphology (specifically, tense) modulates anticipatory eye movements. Importantly, the authors argue that this finding also shows that the relationship between eye-movements and language processing cannot be described as simply one of establishing (or anticipating) reference: The empty wine glass could have not possibly been an upcoming referent of the verb *drunk*, since it violates the selectional restrictions of the verb (drinking requires a liquid, not an empty glass). What the anticipatory eye movements in this case suggest is some form of reasoning about what the end state of an event of *having had drunk* would look like – namely an empty glass.

Altmann and Kamide (2007) concluded by stating that the guiding force in language-vision interactions is one of **conceptual overlap** and not of establishing reference: The conceptual features retrieved from a visual representation are matched to the semantic features retrieved from the linguistic input. The greater the overlap, the greater the likelihood that participants will fixate that specific visual representation.

This linking hypothesis is compatible with an interpretation of the results of the vehicle region that states that the word "princess" undergoes spontaneous lexical modulation given the parameters provided by context and the verbal topic, as stated by the Category Inclusion view: The conceptual overlap that participants calculate when allocating their visual resources is not one between the picture "cat" and the word *princess* but between the picture "cat" and PRINCESS*, an ad hoc category created to describe things that are spoiled. In the metaphoric late-verb conditions, the parameters for modulation are not narrow enough to trigger the modulation of "princess" because the verbal topic has not been heard yet. Participants therefore look preferentially at the literal picture when hearing the vehicle in this condition because the word "princess" has not been modulated yet: Participants established the biggest conceptual overlap between the word "princess" and the picture of a princess. Altmann and Kamide (2007)'s linking hypothesis can therefore account for the present results when assuming a Category Inclusion view.

The Implicit Comparison View would struggle to accommodate the results of the vehicle region in the metaphoric conditions under the linking hypothesis of Altmann and Kamide (2007): If the concept of *princess* is retrieved and aligned with the verbal topic, the initial conceptual overlap of *princess* and the visual representation of a literal princess should be the same regardless of position in the sentence. As Experiment 5 shows, this does not appear to be the case.

A crucial part of this interpretation is the fact that, in the vehicle region, the ratio of looks to metaphoric over literal picture was not modulated by verb position in the literal conditions. This suggests that given a strong linguistic context biasing participants in one direction (Sebastian has a friend who is a noble woman and is hungry) a verb's thematic role constraints can be bypassed to some degree, resulting in the word "princess" eliciting the same referential effect (or in the terms of Altmann and Kamide (2007), the same degree of conceptual overlap between vehicle and literal picture) regardless of whether the word is heard before or after the verb *feeds*: In both cases, the conceptual overlap between linguistic and visual stimuli is similarly large. This attributes a pivotal role to the linguistic context: It can both bypass a verb's thematic role constraints (in the literal conditions) and work together with them to enable the interpretation of "princess" as a metaphor (in the metaphoric conditions).

What happens then, in the absence of a linguistic context? Metaphors in general and the metaphors used in this study in particular, are extremely context-dependent: Take our example sentence *Sebastian feeds a princess*: Without contextual information about Sebastian having a spoiled cat, it should be very difficult to infer that *princess* is a metaphor for "spoiled". We would therefore expect participants to construct a literal interpretation of *princess* regardless of verb position. On the other hand, without a linguistic context, participants might rely more strongly on the thematic role constraints of the verbs: After hearing *füttert* they might anticipate a cat as the most likely referent first and experience confusion when hearing the word *princess*. This could result in less looks to the image of the literal princess when hearing the word "princess" in the early-verb compared to the late-verb condition. If this is the case, this would suggest that the results of Experiment 5 might be a consequence of the verbs' thematic role constraints and not of the metaphor comprehension process.

This reasoning sets the stage for a follow-up study to test how the utterances of Experiment 5 are understood in the absence of a linguistic context. This follow-up is presented below as Experiment 6.

## 4.2 Experiment 6

Experiment 5 suggests that there are processing differences when understanding a verb-object metaphor as a function of the relative syntactic position of the metaphor's elements. On the other hand, when the same verb-object pair is interpreted literally, their relative syntactic position does not seem to affect the way in which the appropriate visual referent is selected when hearing the object of the sentence. This difference between literal and metaphoric conditions suggests that the construction of metaphoric meaning is **asymmetric**, contingent upon the order of presentation of the metaphor's elements, whereas no such asymmetry is visible when participants constructed a literal interpretation.

Crucially, this difference is brought on by the interpretative constraints introduced by the written linguistic context, which guided participants unequivocally towards a literal or a metaphoric interpretation of the heard utterances. What would participants' viewing behavior look like in the absence of the linguistic context? Would participants default to a literal interpretation of the heard utterance (resulting in a viewing pattern similar to that found in the literal conditions of Experiment 5) or will their viewing pattern be contingent upon sentence structure (resulting in a viewing pattern similar to that found in the metaphoric conditions of Experiment 5)?

Experiment 5 also showed that a linguistic context and a verb's selectional restrictions work together to generate expectations about upcoming verb objects: When the two were aligned (that is, when both linguistic context and the verb's selectional restrictions were compatible with anticipating the same visual referent, as in the metaphoric early-verb condition), anticipation of an upcoming object took place: In the metaphoric early-verb condition, participants preferentially looked at the picture of the cat when hearing the verb. When the two sources of information conflicted (as in the literal early-verb condition, where context suggested the picture of the princess as referent but the verb's selectional restrictions suggested the cat), participants did not anticipate a specific upcoming object: In the literal early-verb condition, participants did not show a preference for looking at neither literal nor metaphoric picture. But what would happen if the contextual anticipation cue was no longer available? It's likely that participants would then anticipate the referent most compatible with the verb's selectional restrictions upon hearing the verb, in line with previous findings (Altmann & Kamide, 1999; Hintz et al., 2017). However, if in the absence of context participants did not anticipate any object whatsoever for the specific verbs used in Experiment 5, it would mean that the results from Experiment 5 would have to be re-evaluated, since the interpretation of these results depends on the assumption that the verb is providing a set of parameters that allowed participants to anticipate the upcoming object. Experiment 6 addresses this possibility by examining the way in which participants' viewing patterns is modulated by sentence structure independently of context. To do this, I created a new version of the previous experiment that is identical to experiment 5 but crucially excluded the presentation of the linguistic context.

### 4.2.1 Participants

Thirty-two native speakers of German (aged 18 to 31) with normal or corrected-to-normal vision gave their informed consent and received 11 Euros each for their participation. None of them participated in any of the previous experiments. All participants were right handed and had normal or corrected-to-normal vision. This study was covered by the ethics vote of the psycholinguistics lab of the Humboldt-Universität zu Berlin.

### 4.2.2 Materials and design

Materials were identical to those used in Experiment 5 with the exception of the written linguistic context. The design consisted therefore of only one factor (verb constraint) with two levels (early and late) for the vehicle, verb, "und", and "adv" regions. The model for the "dis" region, however, included an additional factor in the analysis ("Disambiguation" with the levels "metaphoric" and "literal"). This is because, in this region, the utterance disambiguated the interpretation of the verb-object pair (*füttert eine Prinzessin*) by explicitly referring to either "the cat" (*die Katze*, metaphoric conditions) or "the noble woman" (*die Adlige*, literal conditions).

### 4.2.3 Predictions

Experiment 5 showed that context is fundamental for understanding verb-object metaphors: The same utterance, preceded by either a literally-biasing or a metaphorically-biasing context resulted in completely different interpretations judging by the eye-gaze patterns and the picture selection task. Since context is crucial for constructing a metaphoric interpretation (as discussed in chapter two), the absence of context should lead participants to default to a literal interpretation of the utterance. If this is the case, I would expect participants in Experiment 6 to draw a literal interpretation of the utterance and subsequently display a viewing pattern similar to that found in the literal conditions in Experiment 5.

Specifically, I would expect participants to consider the literal representation of the metaphoric vehicle (the princess) as an appropriate referent when hearing the word *Prinzessin* in a similar way regardless of sentence structure. Such a finding would strengthen the interpretation of the results of Experiment 5, given that it would suggest that the asymmetries in processing found in that study are a product of the construction of context-biased metaphoric meaning and not just of sentence structure.

If, on the other hand, participants' viewing pattern varies as a function of sentence structure in a similar fashion as seen in the metaphoric conditions of Experiment 5, it would suggest that this asymmetric viewing pattern is not necessarily a consequence of the way a metaphoric interpretation is constructed through context. This finding would weaken the interpretation of the results of Experiment 5, since it would suggest that they do not reflect metaphoric processing exclusively.

Concretely, if differences in the viewing behavior in Experiment 5 between early and late verb conditions are a consequence of contextually-derived metaphoric interpretation, we should expect to find no overall difference in log-gaze probability ratios of looks to metaphoric picture over looks to the literal picture between early and late verb conditions in the current, context-free, experiment.

When participants hear the verb prior to the object (*Sebastian feeds the princess*), they have nothing else to rely on in terms of anticipating what the verb's object is going to be besides the verb's selectional preferences. When hearing the verb after the object, it is natural that their visual attention would shift towards the literal depiction of said object ("the princess"), given that this is the linguistic input that they have already encountered (in the sentence: *Sebastian will the princess feed*).

Because of this, the main prediction in this regard is that if participants rely on a verb's selectional preferences to anticipate objects, there should be a significant difference between early and late verb conditions in the verb region, with a larger log-gaze probability in the early compared to the late condition. Furthermore, the log-gaze ratio of the early-verb condition should be positive and significantly different from zero, which would represent an overall preference for the verb's ("feeds") typical object ("the cat") compared to the competitor image ("the princess").

At this point, it is important to consider one key difference between Experiment 5 and Experiment 6. Given that in Experiment 6 there is no context biasing towards either literal or metaphoric interpretation, the literal picture can no longer be considered a true competitor. Instead it is a further distractor: When hearing the verb prior to the object (early-verb condition), participants have no reason to consider the princess as a viable referent. This is why, for the verb region, two additional analyses were carried out using the log-gaze probabilities of looks to metaphoric picture over looks to distractor image 1 and distractor image 2 as dependent measures. By doing this, it should be possible to find that participants preferred looking at the target picture (cat) when hearing the verb (*feeds*) compared to all other images on the display.

For the purpose of completeness, I also analyzed the following 3 regions ("und", "adv" and "dis") just as it was done for Experiment 5. The results of the analyses for these last three regions are reported in Appendix D. No differences between verb conditions were predicted for any of these regions. However, in the DIS region, there should be an effect of the disambiguating word: Participants should look at the image that matches the word they are hearing.

### 4.2.4 Procedure

The procedure was identical to that of Experiment 5, with the exception of there not being a written context prior to the target utterance. Every trial therefore began with the two-second preview of the visual grid.

### 4.2.5 Analysis

Experiment 6 was analyzed in a similar way as Experiment 5: Models including verb constraint as a fixed effect as well as trial order as a control were fitted to the log-gaze probability ratios of looks to metaphoric picture over looks to literal picture in the vehicle and verb regions. All models used a treatment contrast coding scheme. They were each fitted twice, changing the intercept of the model from the early to the late verb constraint condition respectively.

Additionally, two further models were fitted to the verb region with different dependent variables: One model used log-gaze probabilities of looks to metaphoric picture divided by looks to the distractor image 1 and the other model used log-gaze probabilities of looks to metaphoric picture divided by looks to the distractor image 2.

### 4.2.6 Results

#### Non-gaze Data

**Picture Selection** Overall, participants preferentially selected the literal picture as the referent at the end of each trial. They chose this image over 90% of the times in the literal conditions and around 45% of the times in the metaphoric conditions. This suggests that without a context, it was difficult to determine whether the utterance was meant metaphorically or not, even after hearing the disambiguating word.

Additionally, participants were faster at selecting an image in the literal compared to the metaphoric conditions. These results can be visualized in Figures 4.11 and 4.12.

#### Vehicle region

There was no significant difference between early and late verb conditions. This result suggests that the difference found in Experiment 5 was caused by the way in which participants reached a contextually-driven metaphoric interpretation and not by structural properties of the sentences when deprived of a context.

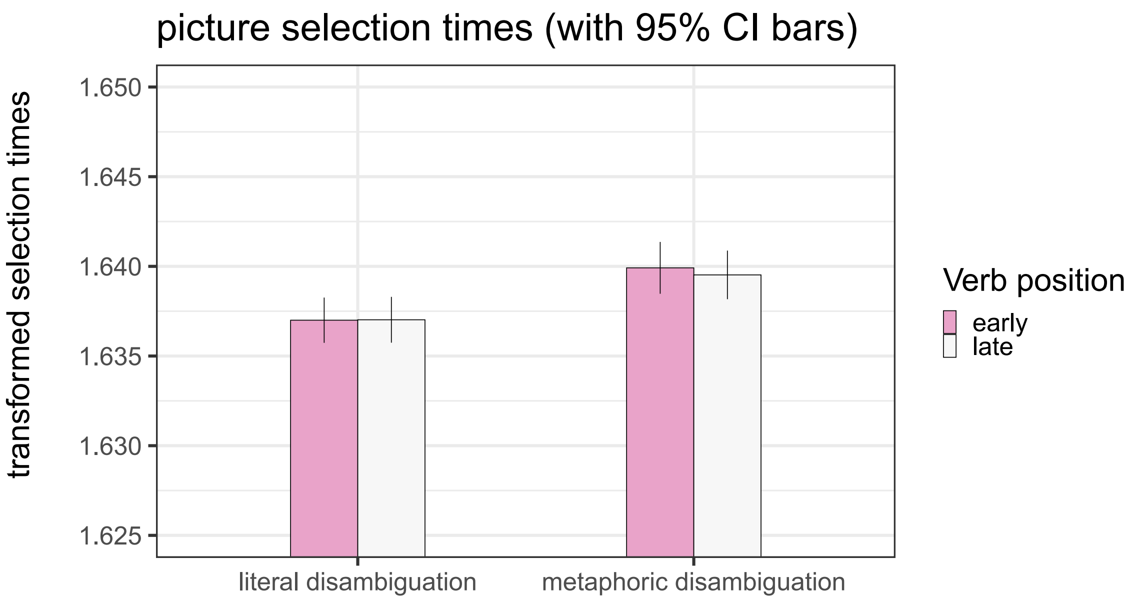


**Figure 4.11:** Picture Selection times for Experiment 6

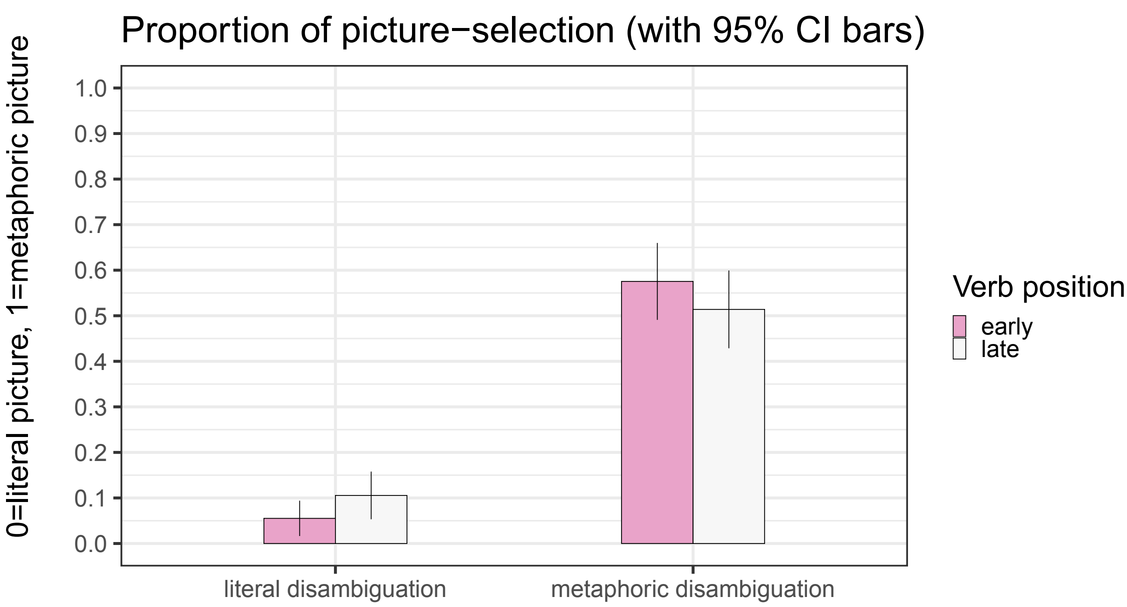


**Figure 4.12:** Proportion of picture selected in Experiment 6. 0 and 1 represent selection of the literal and metaphoric picture respectively.

Because this interpretation is based on a null-effect, it was necessary to show that Experiment 6 was sufficiently powered to detect an effect if there had been one. For this purpose, a power curve was computed through simulations using the output of the model ran for Experiment 5 for this same time-window. That model had shown an effect of verb position in the metaphoric conditions, with a larger log-gaze ratio in the early compared to the late conditions (see Table 4.4). For the power curve simulations, I assumed a true effect size half of the one found in Experiment 5, following Green and MacLeod (2016), in order to have a conservative estimate of the number of participants needed to detect a small effect size. The simulation showed that an experiment with 22 participants would have over 80% power to detect an effect half as big as the one found in Experiment 5. This suggests that Experiment 6 was more than sufficiently powered to find a small difference between verb conditions and yet failed to find one. This strengthens the claim that the difference between verb conditions found in Experiment 5 was indeed a consequence of the interpretation biases provided by the linguistic context, suggesting that it was the way the metaphors were processed in context and not the differences in the sentence structure per se that brought about the effect found in Experiment 5.

Furthermore, participants displayed a preference for fixating the literal compared to the metaphoric picture throughout the region. This is evidenced by the negative intercept of both versions of the model, which were significantly different from zero. Figure 4.13 illustrates the results for the vehicle region of Experiment 6, and Table 4.9 shows the output of the statistical model.

### Verb Region

The first model fitted in this region (log-gaze ratio of looks to metaphoric over looks to literal picture as dependent measure) showed that there was a significant difference between early and late-verb conditions, with a larger log-gaze ratio in the early compared to the late condition. This means that participants looked more at the metaphoric picture compared to the literal picture when they heard the verb before compared to after the object. The second and third model, which compared looks to the metaphoric picture to looks to the distractor images, showed a different result: Here, there was no difference in log-ratios between the verb conditions.

The first model also showed a positive intercept that was not significantly different from 0. This means that there was no overall preference for looking to the metaphoric picture compared to the literal picture in the early-verb condition in this region. This can be visualized in Figure 4.14, and Table 4.10 shows the output of the model. It is also visible, however, that a preference for the metaphoric picture arises as time in this region progresses. In fact, when refitting this model on the second half of this region only, the intercept of the model is positive and significantly different from zero. Table 4.11 shows the output of this refitted model. This suggests that participants are able to anticipate the metaphoric picture when hearing the verb prior to the object, but the effect only arises later in time.

The second and third models both showed positive intercepts that were significantly different from zero. This means that throughout this entire region, participants looked more to the metaphoric picture compared to either one of the distractor images when hearing the verb prior to the object. This is shown in Figures 4.15 and 4.16. The output of the models can be seen in Tables 4.12 and 4.13.

Taken together, these results suggest that in the absence of a linguistic context participants mostly relied on their knowledge of a verb's selectional preferences in order to anticipate what they believed to be the most likely upcoming referent. However, the results from the first model (Table 4.10) suggest that this might occur with somewhat of a delay.

### 4.2.7 Discussion

Experiment 6 was primarily intended as a baseline to assess the results of Experiment 5: The two Experiments were identical with the exception of the linguistic context, which was absent in Experiment 6. The results of the vehicle region of Experiment 6 showed no difference between early and late verb constraint conditions. This suggests that the effect of verb constraint found in Experiment 5 was indeed caused by the biasing linguistic context and not by the structure of the sentence.

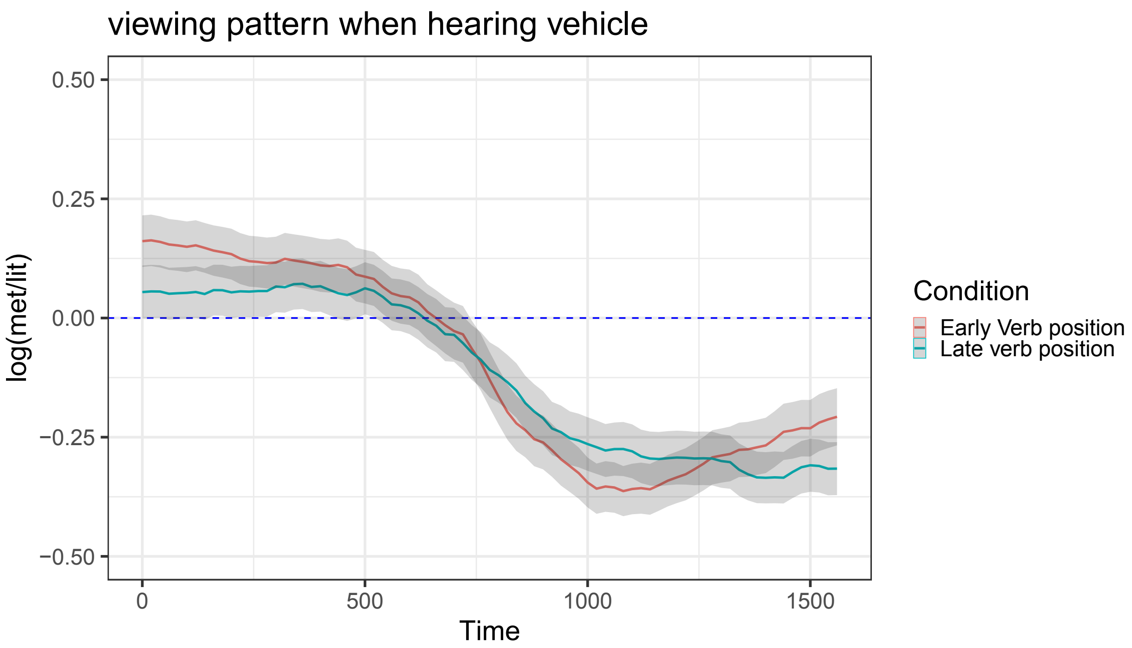


**Figure 4.13:** Summary of results for the vehicle region, Experiment 6.

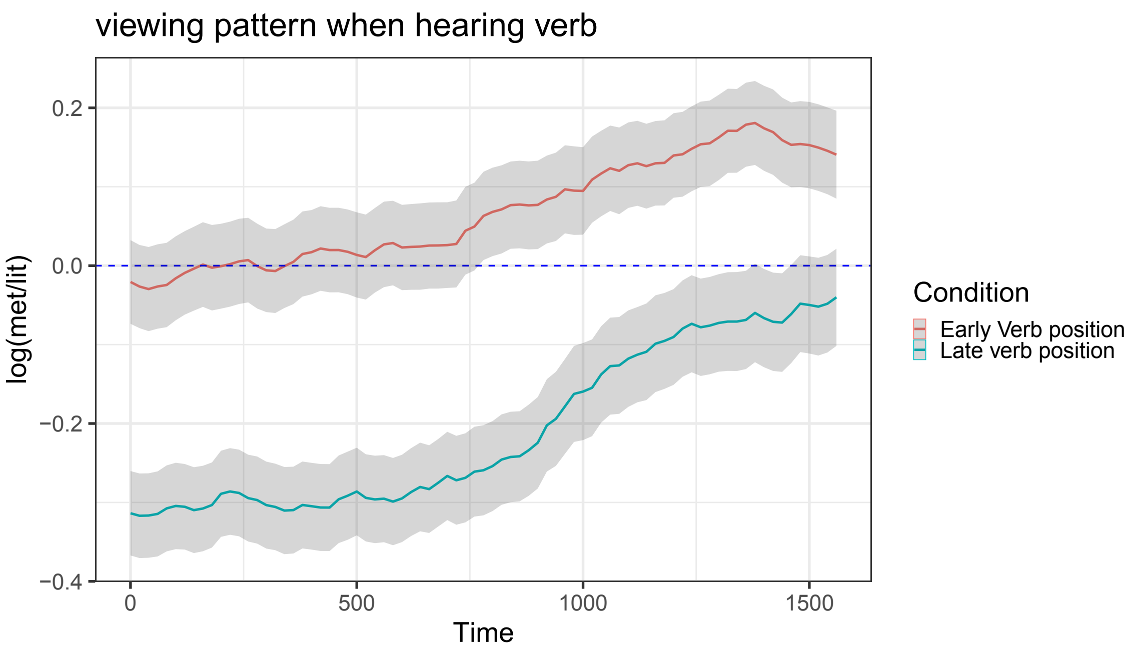


**Figure 4.14:** Summary of results for the verb region, Experiment 6.

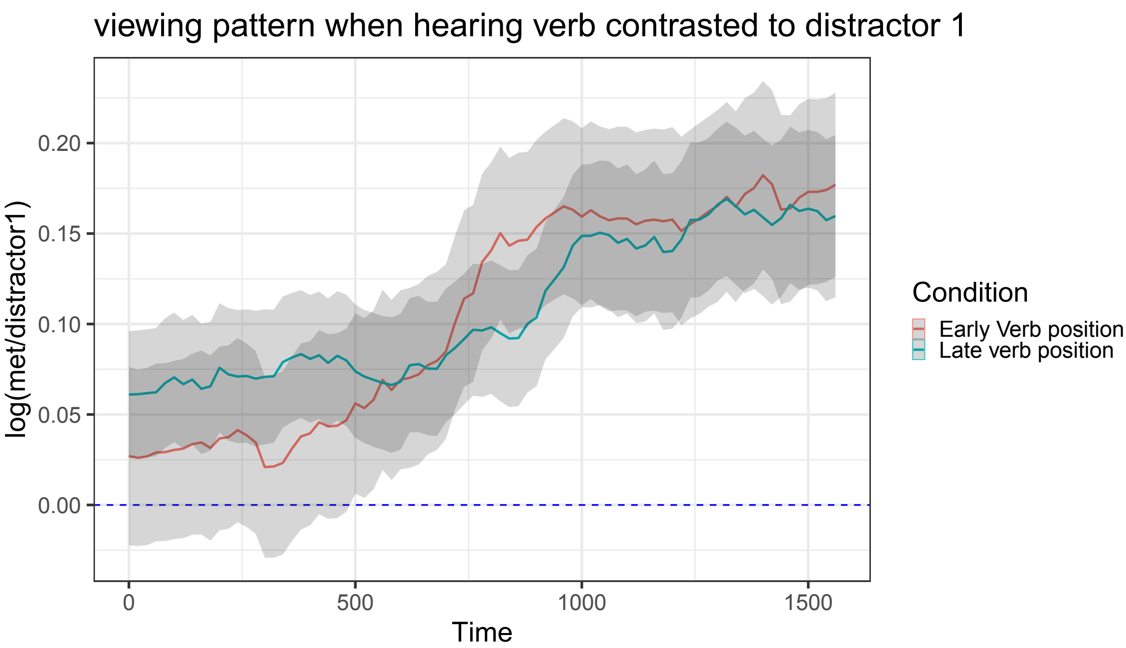


**Figure 4.15:** Summary of results for the verb region contrasted to distractor 1, Experiment 6.

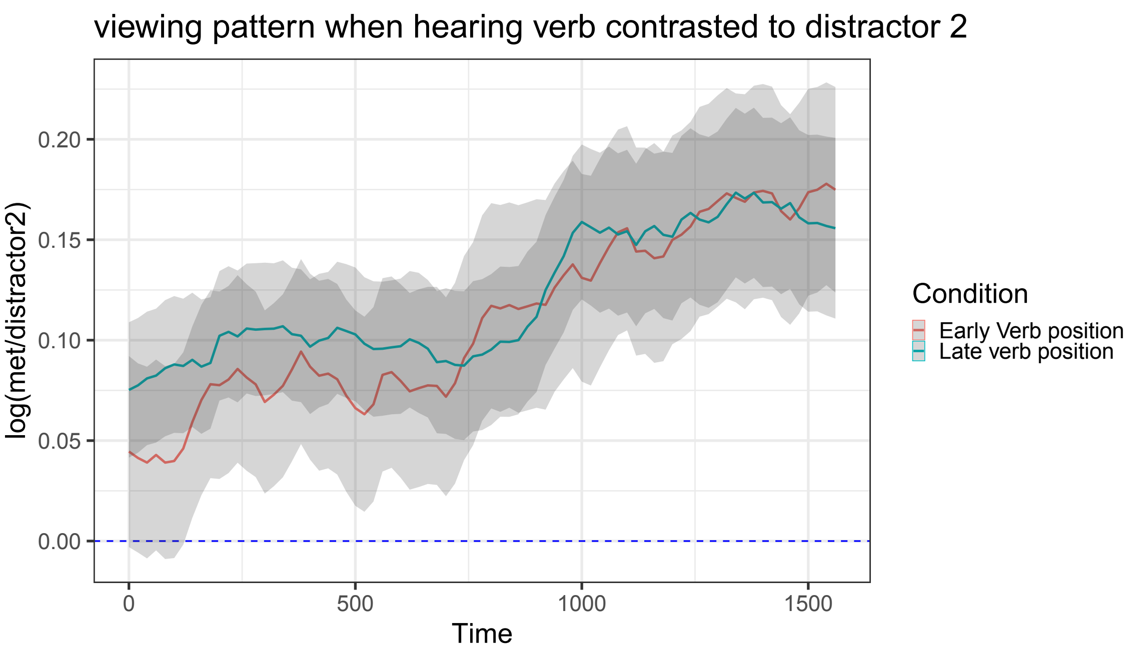


**Figure 4.16:** Summary of results for the verb region contrasted to distractor 2, Experiment 6.

**Table 4.9:** Regression analysis of log-gaze probability ratio in the vehicle region of Experiment 6

| | *Dependent variable:* | |
|---|---|---|
| | logratio | |
| | (1) | (2) |
| Early-Metaphoric vs. Early-Literal | 0.019<br>(−0.096, 0.135)<br>t = 0.331 | |
| Trial Order | 0.002<br>(0.002, 0.003)<br>t = 9.371*** | 0.002<br>(0.002, 0.003)<br>t = 9.371*** |
| Intercept | −0.160<br>(−0.227, −0.093)<br>t = −4.677*** | −0.140<br>(−0.213, −0.068)<br>t = −3.808*** |
| condition at intercept | early-met | late-met |
| Observations | 55,300 | 55,300 |
| Log Likelihood | −44,472.370 | −44,472.370 |
| Akaike Inf. Crit. | 88,960.750 | 88,960.750 |
| Bayesian Inf. Crit. | 89,032.110 | 89,032.110 |

*Note:* *p<0.05; **p<0.01; ***p<0.001
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

Besides serving as a control for Experiment 5, there are some other interesting conclusions that can be drawn from Experiment 6. The results of the non-gaze data, for example, show that in the absence of a context, verb-object metaphors are hardly understood: Participants preferentially selected the literal picture at the end of the trial and took longer to select the metaphoric picture compared to the literal picture, despite having already heard the disambiguating word at the time of selection. This suggests that the selectional preferences of the verb (*füttern* prefers to select an animal as an object over an adult human) were on average not strong enough to force a metaphoric interpretation of the verb's object. Instead, participants seemed to have loosened the verb's selectional restrictions and understood the object literally (i.e., feeding a literal princess).

**Table 4.10:** Regression analysis of log-gaze probability ratio in the verb region of Experiment 6

| | *Dependent variable:* | |
|---|---|---|
| | logratio | |
| | (1) | (2) |
| Early-Metaphoric vs. Early-Literal | −0.271 | |
| | (−0.383, −0.158) | |
| | t = −4.720*** | |
| Trial Order | 0.002 | 0.002 |
| | (0.001, 0.002) | (0.001, 0.002) |
| | t = 8.306*** | t = 8.306*** |
| Intercept | 0.031 | −0.239 |
| | (−0.045, 0.107) | (−0.303, −0.176) |
| | t = 0.805 | t = −7.411*** |
| condition at intercept | early-met | late-met |
| Observations | 55,774 | 55,774 |
| Log Likelihood | −42,005.740 | −42,005.740 |
| Akaike Inf. Crit. | 84,027.480 | 84,027.480 |
| Bayesian Inf. Crit. | 84,098.910 | 84,098.910 |
| *Note:* | *p<0.05; **p<0.01; ***p<0.001 | |
| | values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) | |

This interpretation is supported by the eye-tracking results. During the vehicle region, participants preferentially directed their gaze to the literal picture over the metaphoric picture in both conditions, judging by the negative log-gaze ratios in both versions of model 4.9, which were significantly different from zero. They did this despite the fact that the eye-tracking record showed a sensitivity to the verb's selectional preferences when hearing the verb: When the verb was heard before the object, participants looked at the picture most associated with the verb's thematic role constraints (the metaphoric picture) more than they looked at any of the two distractors (throughout the entire region, Tables 4.12, 4.13) and more than they looked at the literal picture (only in the second half of the verb region, Table 4.11).

Overall, these results are consistent with what has been dubbed *the verb*

**Table 4.11:** Regression analysis of log-gaze probability ratio in the second half of the verb region of Experiment 6

| | *Dependent variable:* | |
|---|---|---|
| | logratio | |
| | (1) | (2) |
| Early-Metaphoric vs. Early-Literal | −0.241 | |
| | (−0.367, −0.116) | |
| | t = −3.759*** | |
| Trial Order | 0.002 | 0.002 |
| | (0.002, 0.003) | (0.002, 0.003) |
| | t = 7.043*** | t = 7.043*** |
| Intercept | 0.083 | −0.159 |
| | (0.003, 0.162) | (−0.239, −0.079) |
| | t = 2.030* | t = −3.903*** |
| condition at intercept | early-met | late-met |
| Observations | 28,240 | 28,240 |
| Log Likelihood | −21,182.630 | −21,182.630 |
| Akaike Inf. Crit. | 42,381.260 | 42,381.260 |
| Bayesian Inf. Crit. | 42,447.250 | 42,447.250 |
| *Note:* | *p<0.05; **p<0.01; ***p<0.001 | |
| | values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) | |

*mutability effect* (Gentner & France, 1988; King & Gentner, 2019): When undergoing semantic strain, a verb's semantic features are more likely to be altered than the semantic features of the verb's subject. What this means is that when comprehending an unusual noun-verb combination (such as *the lizard worshiped*) people are more likely to adjust the verb's meaning to match the features of the noun than the other way around. Gentner and France (1988) investigated this with two Experiments that showed that participants preferred to paraphrases sentences such as *the lizard worshiped the sun* as *The reptile stared unblinkingly at the sun*, leaving the meaning of the subject noun intact and adjusting the meaning of the verb (instead of, for example, interpreting the noun *lizard* as a priest and keeping the meaning of the verb *worshiped* intact). Experiment 6 of the current investigation extends these

**Table 4.12:** Regression analysis of log-gaze probability ratio in the verb region of Experiment 6. DV uses looks to distractor 1

| | *Dependent variable:* | |
|---|---|---|
| | logratio1 | |
| | (1) | (2) |
| Early-Metaphoric vs. Early-Literal | 0.028<br>(−0.055, 0.110)<br>t = 0.657 | |
| Trial Order | 0.001<br>(0.001, 0.002)<br>t = 6.965*** | 0.001<br>(0.001, 0.002)<br>t = 6.965*** |
| Intercept | 0.064<br>(0.007, 0.121)<br>t = 2.191* | 0.092<br>(0.034, 0.150)<br>t = 3.098** |
| condition at intercept | early-met | late-met |
| Observations | 55,774 | 55,774 |
| Log Likelihood | −29,931.400 | −29,931.400 |
| Akaike Inf. Crit. | 59,878.790 | 59,878.790 |
| Bayesian Inf. Crit. | 59,950.220 | 59,950.220 |
| *Note:* | *p<0.05; **p<0.01; ***p<0.001<br>values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) | |

findings to show that the verb mutability effect also seems to apply for verb-object combinations such as *füttert eine Prinzessin*.

The fact that participants on average did not anticipate the metaphoric picture over the literal picture during the first half of the verb region could have two possible explanations.

First, it could be the case that participants developed a task-specific behavior throughout the experiment: Since for all critical items the object of the verb was never the one most likely to be associated with the verb, participants might have adopted the strategy to ignore the information coming from the verb's selectional preferences. What speaks against this is the fact that trial order had a significant effect throughout the region: As the experiment progressed, participants on average

**Table 4.13:** Regression analysis of log-gaze probability ratio in the verb region of Experiment 6. DV uses looks to distractor 2

| | *Dependent variable:* | |
|---|---|---|
| | logratio2 | |
| | (1) | (2) |
| Early-Metaphoric vs. Early-Literal | 0.020<br>(−0.050, 0.090)<br>t = 0.548 | |
| Trial Order | 0.001<br>(0.0003, 0.001)<br>t = 3.596*** | 0.001<br>(0.0003, 0.001)<br>t = 3.710*** |
| Intercept | 0.093<br>(0.042, 0.143)<br>t = 3.585*** | 0.117<br>(0.068, 0.166)<br>t = 4.649*** |
| condition at intercept | early-met | late-met |
| Observations | 55,774 | 55,774 |
| Log Likelihood | −28,440.170 | −28,655.610 |
| Akaike Inf. Crit. | 56,896.340 | 57,325.210 |
| Bayesian Inf. Crit. | 56,967.770 | 57,387.710 |
| *Note:* | *p<0.05; **p<0.01; ***p<0.001<br>values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) | |

tended to look more at the metaphoric picture over the literal one. Why would they look *more* at the metaphoric picture throughout the experiment if they had developed a strategy to ignore the verb's selectional preferences?

A second explanation is simply that anticipation effects based solely on a verbs selectional restrictions are somewhat delayed in comparison to anticipation effects based on a combination of selectional preferences and contextual biases: In the verb region of Experiment 5, there was an overall preference for anticipating the metaphoric picture over the literal one for the entire duration of the region (see Table 4.5), an effect that visibly appears somewhere around 500 milliseconds after verb onset. In Experiment 6, this effect visibly appears at around 750 milliseconds.

## 4.3 Interim Conclusion

Taken together, the results of Experiments 5 and 6 suggest that the processing of verb-object metaphors is asymmetric: When the verb appeared before the vehicle, participants preferentially directed their gaze towards the image depicting a metaphoric interpretation when hearing the vehicle. When they heard the vehicle prior to the verb, on the other hand, participants preferred inspecting the image that depicted a literal interpretation of the vehicle. In the absence of a linguistic context (Experiment 6) no such difference in processing was found. This points to a difference in the way in which the metaphoric meaning was constructed as a function of the position of the elements, in line with the Category Inclusion View (Glucksberg, 2001; Glucksberg et al., 1997) and at odds with the Implicit Comparison view (Wolff & Gentner, 2000; Wolff & Gentner, 2011).

As previously discussed, this interpretation of the results is compatible with the conceptual overlap linking hypothesis (Altmann & Kamide, 2007): When hearing the word *Prinzessin*, participants evaluated the degree of conceptual overlap with the available visual context: In the metaphoric early-verb condition, they deemed the overlap to be greatest with the metaphoric picture and in the metaphoric late-verb condition they identified the literal picture as having the greatest overlap. This suggests that the semantic features of the word *Prinzessin* varied as a function of its position in the sentence: When appearing before the verb, it likely retained most of the semantic features associated with a literal interpretation of the word (hence the overlap with the literal picture). When appearing after the verb, the semantic features associated with the word changed and participants likely constructed the ad hoc category PRINZESSIN*, which has a greater conceptual overlap with the metaphoric picture (the cat).

Importantly, there were no differences in picture selection accuracy or speed in Experiment 5. In Experiment 5, participants also settled on the metaphoric picture as the most likely referent in both metaphoric conditions from the region “und” onward (judging by the gaze record). This suggests that despite the early asymmetries in processing, participants still derived a metaphoric interpretation in each case with similar processing ease and accuracy.

In terms of what the results of these experiments say about the way in which an object is anticipated given the verb, there are two conclusions to be drawn. Firstly, it seems that even though a verb's selectional restrictions influence participants' eye-gaze, they are not strong enough to override the mental representations derived from a previous linguistic context. In Experiment 5, participants showed sensitivity to the verb's selectional restrictions in the early verb-metaphoric condition, where they quickly anticipated the referent that was compatible with both the context and the verb. However, when context was pitted against the verb's selectional preferences, participants did not anticipate any referent and instead seem to wait for further linguistic input before settling on an interpretation.

Secondly, participants are more likely to loosen a verb's selectional preferences than to loosen the semantic features of the verb's object when these collide. Participants preferred to derive literal interpretations (and did so more quickly) of sentences such as *Sebastian feeds a princess*, indicating that they adjusted the verb's but not the noun's semantic properties. This is analogous to findings showing that verb's will be more likely loosened compared to their sentential subject, which is known as the the verb mutability effect (Gentner & France, 1988).

I now turn to the final empirical test of the two theories discussed throughout this dissertation. As discussed in chapter two, mediating effects of aptness, familiarity and conventionality are theoretically relevant for both Category Inclusion and Implicit Comparison views. In the next chapter, I provide a re-examination of Experiments 2 and 5 in order to explore the relationship between metaphor processing and metaphoric aptness, familiarity and conventionality in the light of the theoretical predictions discussed in chapter two.

# 5

# Assessing the Impact of Aptness, Conventionality and Familiarity on Metaphor Comprehension

This chapter presents the dissertation's last empirical test of the discussed theories of metaphor comprehension. Here, I address the following question:

- *What is the relative contribution of aptness, conventionality and familiarity to the processing of metaphors in Experiments 2 and 5?*

This question is important to ask given the theoretical repercussions it might have: It is namely the case that these mediating factors are ascribed different functions in the two main theories discussed in this dissertation.

Bowdle and Gentner (2005) formulated the "career of metaphor hypothesis" as an addendum to Gentner's Structure Mapping Theory. It suggests that novel metaphors are processed via analogical reasoning, but, the more often a metaphor is heard, i.e. the higher the metaphor's degree of conventionality, the more likely it is that it will be processed as a category statement: When a metaphoric vehicle is applied to different metaphoric topics to yield the same figurative interpretation (by activating the same relational structure patterns), the vehicle will begin to be conventionally associated with this abstracted, metaphoric meaning and thus become polysemous. Once this has occurred, the expression as a whole will be processed via class inclusion: The vehicle will have acquired a secondary, abstract-categorical meaning as a by-product of repeated structural alignment.

Crucially, the career of metaphor hypothesis sees conventionality impacting the stage of structural alignment: More conventional metaphors will display a *vertical alignment*, i.e. one in which the vehicle is understood as a higher-order category abstraction, with the topic being a member of said category. Less conventional metaphors, on the other hand, establish *horizontal alignments* in which both terms are on the same level of abstraction. This means that degree of vehicle conventionality impacts the stage of structural alignment, which according to Structure Mapping Theory, occurs prior to the projection of inferences from vehicle to topic.

Bowdle and Gentner (2005) also provide empirical support for their view. First, they discuss the literature showing how conventional metaphors are understood faster than novel metaphors (Blank, 1988; Gentner & Wolff, 1997). They interpret the reading time advantage for conventional metaphors as evidence for a switch in processing mode. However, the main evidence supporting their view comes from 3 experiments in which they evaluate the effect of conventionality on form preference by comparing the processing of similes and metaphors with identical conceptual information (e.g. *A soldier is a pawn* vs. *A soldier is like a pawn*). They rely on the following auxiliary hypothesis: Similes should always be processed as comparisons by default (since they include the explicit comparison trigger *like*), whereas metaphors could preferentially be processed as categorical statements (since they display the surface form of a literal category statement such as *an apple is a fruit*).

In the first two Experiments, the authors showed that conventional metaphors are preferred and read faster in the metaphor form whereas novel ones are preferred and read faster in the simile form. In a third experiment, they conducted a form of "in vitro" conventionalization: By showing participants the same metaphoric vehicle used together with different topics they managed to shift participants' surface form preference from simile to metaphor when the vehicle was seen again, effectively accelerating the conventionalization process in the lab.

Category inclusion theorists have disputed the career of metaphor hypothesis and the evidence supporting it. The argument is best summed up in what Glucksberg (2008) informally deemed the "quality-of-metaphor hypothesis": If a metaphoric vehicle can easily be perceived as capturing important salient features of the topic, it will be modulated into an ad hoc category and the metaphor as a whole will swiftly be understood as a category statement. If, on the other hand, it is not possible to understand the way in which the vehicle captures such salient relevant properties, vehicle and topic will likely be scanned for structural similarities, forcing the metaphor to be processed via indirect comparison. With this in mind, it would not be a metaphor's conventionality value that plays the key mediating role, but a metaphor's aptness: The degree to which the vehicle captures salient relevant properties of the topic.

The strongest case for the "quality-of-metaphor hypothesis" was made by Jones and Estes (2006). They argue against the results of Bowdle and Gentner (2005) by pointing out that in that study the authors had not controlled for the effect of aptness. Aptness and conventionality have usually been found to be highly correlated: Studies showing an effect of aptness on metaphor comprehension did not control for conventionality (e.g. D. Chiappe, Kennedy, & Smykowski, 2003; Gagné, 2002; Jones & Estes, 2005; Kusumi, 1987) and studies showing effects of conventionality did not control for aptness (Bowdle & Gentner, 2005; Gentner & Wolff, 1997). Jones and Estes (2006) created a new set of materials in which conventionality and aptness were carefully teased apart, with an equal number of high-apt/low-conventionality, low-apt/low-conventionality, low-apt/high-conventionality, and high-apt/high-conventionality items. they re-conducted Experiment 2 of Bowdle and Gentner (2005) and found that only aptness predicted form preference and reading times of this set of metaphors and not conventionality. They conclude that aptness is the true mediator of processing form, providing support for the "quality-of-metaphor" hypothesis.

Thibodeau and Durgin (2011) argued against the conclusions drawn by Jones and Estes (2006) on the basis of the definition of conventionality. They argue that Jones and Estes' definition is inaccurate and claim that conventionality should be measured as a sentence-level property and not as a word-level property. This means that a metaphor's conventionality score should not be computed by examining the metaphoric vehicle alone, but by observing the frequency of topic-vehicle pairings, a construct also referred to as metaphor familiarity (Blasko & Connine, 1993; Bowdle & Gentner, 2005). They believe that familiarity and not aptness is responsible for modulating the processing mode of metaphoric expressions.

Besides this definitional point, Thibodeau and Durgin (2011) make two main claims. The first one concerns the explanatory value of using participants' ratings to obtain point estimates of aptness and familiarity and the second one the sensitivity of aptness to context. It is important for the current purposes to have a closer look at these two claims.

Regarding their first claim, they argue that participants might be inadvertently rating familiarity when asked to rate aptness or vice versa. More importantly, they argue that both of these measurements might actually be measuring a third construct known as processing fluency, or how easy it is to process a sentence. Previous work in the domain of social psychology has shown that participants tend to misattribute a broader sense of processing fluency to various different abstract categories of a sentence they are asked to rate (Alter & Oppenheimer, 2009; Jacoby et al., 1988; Jacoby & Whitehouse, 1989). Thibodeau and Durgin (2011) use this as an argument against the validity of the claims regarding aptness made by Jones and Estes (2006), by stating that Jones and Estes (2006) weren't really measuring aptness but processing fluency, so their results cannot provide actual support for the quality-of-metaphor hypothesis. This claim is hard to address because it is not backed by an independent test of the degree to which processing fluency affects ratings of aptness. More importantly, if it is the case that processing fluency affects aptness ratings, there is no reason to believe it shouldn't affect ratings for conventionality and/or familiarity in a similar fashion, meaning that the effect of processing fluency could probably habe been constant across all three dimensions in previous studies.

The second claim, concerns the context-sensitivity of aptness. Thibodeau and Durgin (2011) state that the definition of aptness as "the degree to which the figurative meaning of the vehicle describes a relevant feature of the topic" (Jones & Estes, 2006, p. 19) means that aptness should not be sensitive to contextual manipulations, given that it is a property of the relationship between vehicle and topic. To test this, they conducted an experiment to see whether aptness ratings were sensitive to different priming conditions that involved context: If aptness ratings were shown to be influenced by a contextual manipulation, it would prove that aptness ratings were not measuring the intended dimension, but were instead only indexing overall processing fluency (which is believed to be context-dependent). In their experiment, participants first read metaphoric primes in which the vehicle either shared the same intended meaning as the target (similar-sense prime: *An Encyclopedia is a lantern*, where *lantern* stands for metaphorically lighting the darkness of ignorance), or had a different intended meaning (alternative-sense prime: *A flag is a lantern*, where *lantern* stands for a signaling device). Participants then rated the target metaphor (Target: *Education is a lantern*, meaning that education shines a light on the darkness of ignorance) as being more apt following the same-sense compared to when it followed the alternative-sense metaphor. The authors claim that because the ratings were sensitive to the presence of salient incompatible properties (whether the prime sentence used the vehicle with the same metaphoric meaning or a different one) they were actually measuring processing fluency and not aptness, which should depend only on the topic/vehicle interaction.

To understand why this conclusion is not entirely licensed, it is worth revisiting the discussion of section 2.2.2 on the role of context for establishing the relevant dimensions for attribution. There, I argued that the interaction between topic and vehicle has to necessarily consider the influence of contextually-derived biases for metaphor interpretation. Examples (9a) and (9b) of chapter two, reproduced below as examples (1) and (2) illustrate this point:

*1) My lawyer always eats more than he should and now he is very overweight. My lawyer is an elephant.*

*2) My lawyer can remember anything you tell him because of his incredibly good memory. My lawyer is an elephant.*

Here, sentences with identical topics trigger a different metaphoric interpretation because of the relevant context. The aptness values of (1) and (2) will therefore be different from one another, given that the contextual biasing of the interpretation (which critically contributes to creating the relevant dimensions for attribution) might be felicitous to different degrees in each case. This point can also be made by looking at the verb-object metaphors studied in chapter four: Without the context, participants were not even able to understand them as metaphors (as evidenced by the results of Experiment 5). That being the case, context can be said to play a fundamental role in determining a metaphor's aptness.

The results of Thibodeau and Durgin (2011) are therefore not surprising. Priming different senses of a metaphoric vehicle will modulate aptness ratings of subsequent metaphors using the same metaphoric vehicle: Participants are evaluating the contextually situated representation of the metaphoric topic in each case and then evaluating its fit with the vehicle: If the context is one about the intellectual triumph over ignorance (similar-sense prime: *An Encyclopedia is a lantern*, where lantern stand for metaphorically lighting the darkness of ignorance), it will be easier to process the metaphor *Education is a lantern* (to the extent that participants consider the sentences to be contextually related). If, on the other hand, the context is a technical discussion on signaling devices (alternative-sense prime: *A flag is* *a lantern*), participants will most likely be uncertain about the meaning of the target and rate the metaphor as being overall less apt.

Ultimately, Thibodeau and Durgin (2011) argue for the need of better estimates for aptness and conventionality in order to determine the way in which they influence metaphor processing. Though I disagree with the arguments they have against the construct validity of aptness, I do agree that it is important to improve on the way in which measures of aptness, conventionality and familiarity (from here on referred to as "the three mediating factors") are measured in order to get closer to answering the question of which one is better suited to explain differences in processing. Concretely, I believe that there are two fronts on which improvement is necessary: Measurement validity and construct specification.

## 5.1 Measurement Validity

In the psychometric literature, measurement validity is defined as the extent to which the way a variable is measured (ratings scales for measuring aptness, for example) is an accurate quantification of the cognitive construct that said variable represents (Price et al., 2015, section 4.2). In the literature on metaphor comprehension, there is consistency in the way in which the aptness, familiarity and conventionality of a metaphor are measured, as can be seen in Table 5.1 below, which presents a brief overview of some of the most popular studies measuring these dimensions. Though this table is not meant to be exhaustive nor comprehensive, it does show that aptness, familiarity and conventionality are (a) almost always measured using a 7-point likert scale in which the data is treated as continuous and not ordinal, and (b) they are often dichotomized (crucially, in the studies important for theory development) in order to create two separate groups of items (e.g., low-apt vs. high apt).

Point (a) is a problem for measurement validity because by treating the data as continuous the measure is distorted and it less-accurately represents the intended construct: If a participant is asked to rate how apt a metaphor is on a scale from 1 (very bad) to 7 (very good), the conceptual distance between 3 and 4 will not

| Study | Variable | Scale | As_continuous | Dichotomized | Type | Language | Participants |
|---|---|---|---|---|---|---|---|
| Bowdle & Gentner (2005) | aptness | likert, 1-10 | yes | yes | nominal | English | 16 |
| Bowdle & Gentner (2005) | conventionality | likert, 1-10 | yes | yes | nominal | English | 16 |
| Blasko and Connine (1993) | aptness | likert, 1-7 | yes | yes | nominal | English | 44 |
| Blasko and Connine (1993) | familiarity | likert, 1-7 | yes | yes | nominal | English | 44 |
| Cardillo et al. (2012) | familiarity | likert, 1-7 | yes | no | nominal | English | 80 |
| Cardillo et al. (2012) | familiarity | likert, 1-7 | yes | no | verbal | English | 80 |
| Chiappe and Kennedy (1999) | aptness | likert, 1-7 | yes | no | nominal | English | 46 |
| Chiappe and Kennedy (2001) | familiarity | likert, 1-7 | yes | no | nominal | English | 16 |
| Chiappe et al. (2003) | aptness | likert, 1-10 | yes | no | nominal | English | 36 |
| Chiappe et al. (2003) | conventionality | likert, 1-7 | yes | no | nominal | English | 66 |
| Gentner and Wolff (1997) | conventionality | likert, 1-7 | yes | yes | nominal | English | 12 |
| Jones and Estes (2006) | aptness | likert, 1-7 | yes | yes | nominal | English | 62 |
| Jones and Estes (2006) | conventionality | likert, 1-7 | yes | yes | nominal | English | 33 |
| Thibodeau and Durgin (2011) | familiarity | likert, 1-7 | yes | yes | nominal | English | 72 |
| Thibodeau et al. (2018) | aptness | likert, 1-7 | yes | no | mixed | English | 1200 |
| Thibodeau et al. (2018) | familiarity | likert, 1-7 | yes | no | mixed | English | 1200 |

**Table 5.1:** Overview of previous studies investigating aptness, familiarity and conventionality

be the same as the distance between 1 and 2. This is a fundamental property of ordinal scales. Continuous scales, by contrast, don't have this property as the distance between individual values is assumed to remain constant. This is why, when describing the central tendency of a set of ordinal data, it is more appropriate to use the median (Stevens, 1946), which is the middle score when scores are ranked in order of magnitude, and is less sensitive to extreme values compared to the mean. For continuous data, on the other hand, the mean is the best measure of central tendency, since the intervals of the scale are considered to have the same conceptual distance. In a recent study, Liddell and Kruschke (2018) make a strong case against treating ordinal data as interval data (i.e. using the mean as the central tendency of ordinal data or fitting a linear regression model on ordinal data) by showing that it leads to a loss of statistical power as well as to an inflated false-positive rate. They specifically address the common practice of averaging out the scores of likert scales in order to treat these averages as a new continuous variable (as is usually done in the metaphor literature, see Table 5.1), by simulating data and performing two different analyses: One taking the averaged data as continuous and one taking the un-averaged data as ordinal (section 7, p. 333). They found that, when there was a true underlying null-effect, the model using the averaged data as a continuous variable greatly overestimated the effect size and showed a significant difference. They also found that, when there was an underlying true effect size, this was detected by the model treating the data as ordinal but not detected by the model using continuous data. In sum, treating averaged ordinal data as continuous inflated both the false positive and the false negative rate.

Point (b) (i.e. the dichotomization of continuous data) also affects measurement validity because it leads to a significant loss of the information carried by the measured variable, which can dramatically distort the relationship between measurement and construct. Imagine if, for example, we wanted to measure how aptness affects reading times. So we conduct an experiment with 20 items, for which we have individually measured their aptness levels. We then group the items into low-apt and high-apt groups by performing a median-split. This hypothetical scenario is depicted in Figure 5.1 below. After we run our experiment, we find that the low-apt group had significantly shorter reading times than the high-apt group (the graph to the left in Figure 5.1). However, for the individual items within each group (the yellow dots in the graph to the right), the trend is actually going in the opposite direction: The higher-apt metaphors are associated with shorter reading times within the low-apt group (i.e. the group of metaphors rated between 0 and 10 for aptness) and also within the high-apt group (i.e. the group of metaphors rated between 10 and 20 for aptness). As the figure shows, if we had analyzed this hypothetical data with aptness as a continuous predictor instead of as a categorical predictor, we would have found that the regression line (depicted in red in the graph to the right) actually went in the opposite direction. By dichotomizing the variables we lost crucial information that made us draw the wrong conclusions from the data.

Though this scenario might not be likely to occur, it does illustrate the dangers of dichotomizing continuous variables (or pseudo-continuous, as is the case of averaged ordinal data). This practice, though common in psychology, has been heavily discussed in fields such as clinical medicine and applied statistics. Here, several

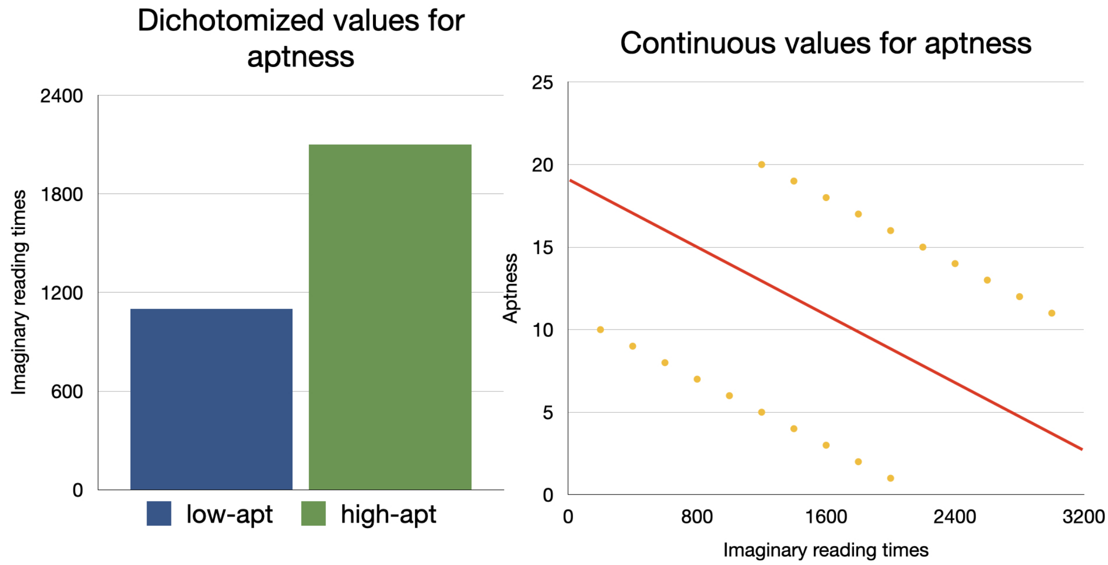


**Figure 5.1:** Hypothetical results of an Experiment on the effect of aptness on reading times.

studies have shown that dichotomizing variables leads to information loss and to a dramatic reduction of statistical power, similar to losing around one third of the data (Altman & Royston, 2006; Cohen, 1983; MacCallum et al., 2002). MacCallum et al. (2002) conducted a series of simulations comparing an analysis using a continuous variables as the predictor in the model vs. using a dichotomized version of the same variable as the predictor. They found that, for the case of a model with a single independent variable, using a dichotomized predictor inflated the false negative rate compared to the model that treated the predictor as continuous. When there were two independent variables in the model, dichotomization lead to an inflated false positive rate. Another simulation study by Austin and Brunner (2004) found that, with increased sample size, the risk of a false positive actually increased with dichotomization of the predictor. Very similar conclusions have been drawn using large real data sets as the test cases (Royston et al., 2006), additionally emphasizing the fact that there is no good reason for selecting the median as an ideal cut-off point. the problem of dichotomization is so severe and so pervasive in many fields of studies that some journals have begun to ban the practice altogether (Dawson & Weiss, 2012).

given this background, it seems likely that points (a) and (b) are affect the validity of the measurements of the three mediating factors in metaphor comprehension. the studies presented in the chapter attempt to improve on this by collecting continuous instead of ordinal data and by not dichotomizing it in the analysis.

## 5.2 Construct Specification

At what point during comprehension do the different theories predict effects of the three mediating factors to appear? Are these effects unique to reading times, or can they be translated to other measurements? Are the effects of the mediating factors predicted for nominal metaphors only, or should they appear in other types of metaphors as well? These questions have not been properly addressed in the literature on metaphor comprehension. As a result, the constructs of aptness, conventionality and familiarity can be said to be theoretically underspecified in regard to these questions.

The career of metaphor appears to be the most explicit in terms of the point in time in which effects should arise, considering that they state the following: "In general, then, conventional metaphors will tend to be interpreted as categorizations rather than as comparisons because the former mode of alignment (i.e. vertical alignment as opposed to horizontal alignment) will be completed more rapidly than the latter" (Bowdle & Gentner, 2005, pg. 199). Considering how the Structure Mapping Engine operates in Structure-Mapping Theory, this means that conventionality should affect the earliest stages of processing, given that this is the point in time in which structural alignment takes place (Wolff & Gentner, 2011). Category Inclusion views, on the other hand, are underspecified in this regard. They claim that aptness should be the key mediating factor, but the point in time in which aptness should play a role remains unclear.

In terms of whether the three mediating factors should affect non-nominal metaphors, the theoretical constructs are also underspecified. The Category Inclusion view states that the mechanisms for processing nominal metaphors also apply to verbal metaphors, such as *The car flew across the intersection* (Glucksberg, 2001, pg. 66-67). It follows from this statement that aptness should also modulate the processing of verbal metaphors, but this is not explicitly stated nor has it been empirically tested. Structure Mapping Theory, on the other hand, does not explicitly state that non-nominal metaphors are processed via analogical reasoning. However, Gentner et al. (2001) discuss the ways in which analogical reasoning can account for the phenomena discussed in Conceptual Metaphor Theory, a theory which does explicitly attempt to account for all types of metaphors with a single explanation. So it it is reasonable to assume that Structure Mapping Theory, in as much as it can capture a wide spectrum of metaphor types, would predict effects of conventionality on non-nominal metaphors. This prediction has also not been previously tested.

Finally, I am not aware of any work considering how the effects on reading times and reaction times found for the three mediating factors in the past (e.g. Blasko & Connine, 1993; Bowdle & Gentner, 2005; Jones & Estes, 2006) should translate to other dependent variables. It is therefore necessary to address these issues in order to better specify the constructs of aptness, conventionality and familiarity.

## 5.3 Mediating Factors in Processing Non-Nominal Metaphors

The experiments presented in this chapter are an attempt to answer the question of which one of the three mediating factors contributes most to metaphor processing. Because the different theories discussed so far make different prediction in this regard, answering this question should provide further insight as to which theory provides the overall best explanation for the phenomenon. To answer this question, I collected new ratings of aptness, conventionality and familiarity for the verbal metaphors used in Experiment 2 and for the verb-object metaphors used in Experiment 5. These ratings were then used to re-fit the models analysis the data from Experiments 2 and 5.

To improve measurement validity compared to previous studies, ratings were collected by using a continuous slider scale ranging from 1-100 instead of an ordinal likert scale. Doing this changes the type of data collected from ordinal to continuous, which makes it possible to take the mean as the measure of central tendency without fearing a loss in statistical power. Additionally, the higher resolution provided by continuous slider scales has been shown to both have a higher validity than coarse-grained likert scales (Warriner et al., 2017) and to produce more reliable results when measuring abstract constructs (Cook et al., 2001; Imbault et al., 2018).

The studies in this chapter also had a larger average sample size than in the majority of previous studies on the mediating factors during metaphor comprehension. By increasing the number of participants, it should be possible to obtain more precise estimates for the aptness, conventionality and familiarity values of each metaphor.

Furthermore, given that the measurements for the three factors have often been found to correlate (Jones & Estes, 2006), I collected ratings from separate group of participants in order to reduce the correlation between the values as much as possible. The ratings were also included in the statistical models as continuous variables, thus avoiding the problems associated with dichotomization discussed in the previous section.

In terms of construct specification, the differences in modality of Experiments 10 (eye-tracking during reading) and 14 (Visual World Paradigm) will allow us to compare how effects vary across dependent measures (reading times, in Experiment 10 and log-gaze ratios, in Experiment 14), and across different types of non-nominal metaphors (verbal metaphors in Experiment 10 and verb-object metaphors in Experiment 14). This will help further specify the constructs of aptness, familiarity and conventionality by examining how they behave across measures and metaphor types. Furthermore, the high sensitivity of the eye-tracking method lets us investigate the time-course of any potential effects of the mediating variables on comprehension. The constructs can therefore be specified with regards to the point in time during processing that their effects should appear.

Overall, the way in which the Experiments in this chapter were constructed should allow us to get better measurements of the three mediating factors and how they impact metaphor processing compared to previous studies. This should in turn allow us to tease apart the effect of the three factors and relate this to the predictions made by the two theories under discussion. I first present the ratings and re-analysis of the verbal metaphors investigated in Experiment 2, followed by the ratings and re-analysis the the verb-object metaphors investigated in Experiment 5.

## 5.4 The Effect of Mediating Factors on Processing Written Verbal Metaphors

Before conducting Experiments 1-4 the metaphors used as critical items were normed for aptness (as described in chapter three). To do this, a group of 20 participants read 40 verbal metaphors together with 40 nonsensical sentences on a scale from 1-7. Based on the results, 4 metaphors were excluded from the main analysis and the remaining 36 were used as the critical items in Experiment 1. An example of a verbal metaphor, including the different analysis regions used in Experiment 2, is presented as sentence (3) below:

(3) *Dass seine / Meinung* NOUN */ umgittert* VERB */ wurde nach dem Regimewechsel, war / schwierig* ADJ */ für den Redakteur*

*"That his / opinion* NOUN */ fenced-in* VERB */ was after the change in regime, was /difficult* ADJ*/ for the journalist"*

"The fact that his opinion was fenced-in after the change in regime was difficult for the journalist."

I used these original ratings to explore their potential effect on reading times of the metaphors in Experiment 2. I chose Experiment 2 given that it was the highest-powered study on verbal metaphors presented in this dissertation. This tentative exploratory study lead me to collect new, higher-quality ratings on aptness as well as on conventionality and familiarity in order to better assess their impact on metaphor comprehension. The collection of these ratings is described in Experiments 7-9. Their individual contribution to the processing of verbal metaphors is explored in Experiment 10.

Since the goal of Experiments 7-9 (as well as 11-13) was the collection of ratings, none of these studies had explicit predictions, and the data collected in these Experiments was only analyzed descriptively.

### 5.4.1 Experiment 7: aptness of verbal metaphors

The first rating study conducted intended to get a new measure of metaphoric aptness for the verbal metaphors used in Experiment 2. This study, as well as all the following rating studies, was programmed and run using an instance of the IBEX farm (Drummond, 2013) coupled with the Penncontroller extension (Zehr & Schwarz, 2018). These tools allow for the creation of web-based experiments that can be run online, without the need of bringing participants to the lab.

#### Participants

50 monolingual German native speakers (ages 18-31) were recruited using the "clickworker" recruiting platform. All participants gave their informed consent prior to taking part in the study and received 2 Euros for their participation. This study was covered by the ethics vote of the psycholinguistics lab of the Humboldt-Universität zu Berlin.

#### Materials and design

The materials used were the same 36 critical items used in Experiment 2. In addition, participants saw 4 ungrammatical, completely non-sensical sentences. These sentences were incorporated as attention controls: Only the data of participants who rated at least 3/4 of these sentences as incomprehensible (i.e. with a value lower than 20 on the provided scale) were analyzed. The presentation of critical and filler items was randomized.

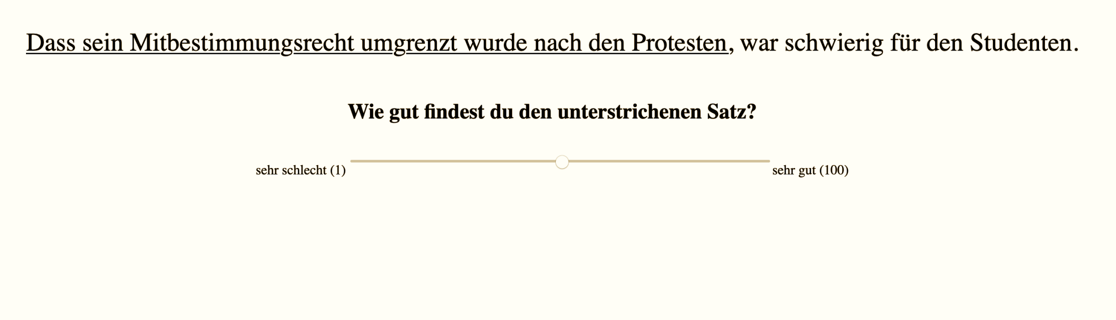


**Figure 5.2:** Example of a trial in Experiment 7

In the study, participants were asked to rate how understandable they thought every metaphor was given the provided context on a continuous, sliding scale from 1 (meaning totally incomprehensible) to 100 (meaning totally comprehensible). The metaphor was always underlined. Participants were not given an explicit definition of a metaphor. This was done in order to avoid biasing their answers as much as possible. An example critical trial can be seen in Figure 5.2.

### Procedure

On each trial, participants first read the entire metaphorical sentence. They then pressed the SPACE bar in order to see the continuous scale. They changed the position of the slider button on the scale using their mouse, after which they clicked on the button "send". The entire Experiment lasted around 10 minutes.

### Analysis and results

Prior to any analysis, the data from 10 participants was excluded based on their ratings of the filler items. This reduced the total number of participants to 40. the data of these 40 participants was used to calculate the mean aptness value of each one of the 36 critical items.

On average, participants rated the items to be somewhat comprehensible, with ratings being shortly below half-way between the ends of the scale (mean value:46.85, median value: 44.81). The lowest rating was 33.20 and the highest was 68.72. The mean aptness values for every item can be seen in Table 5.4.

### 5.4.2 Experiment 8: familiarity of verbal metaphors

The second rating study conducted intended to measure participants' subjective impression of the frequency of each metaphoric expression.

#### Participants

50 monolingual German native speakers (ages 18-31) were recruited using the "clickworker" recruiting platform. None of them participated in Experiment 7. All participants gave their informed consent prior to taking part in the study and received 2 Euros for their participation. This study was approved by the ethics vote of the psycholinguistics lab of the Humboldt-Universität zu Berlin.

#### Materials, design and procedure

The materials, design, predictions and procedure were identical to those of Experiment 7. The only difference was the type of instructions that participants received regarding the task: Instead of being asked to rate how comprehensible the sentences were, the task was to indicate roughly how often they believed to have heard each metaphor on a sliding scale from 1 (never before) to 100 (very often).

#### Analysis and results

Prior to any analysis, the data from 14 participants was excluded based on their ratings of the filler items. This reduced the total number of participants to 36. Their data was used to calculate the mean familiarity value of each one of the 36 critical items. Only descriptive statistics were obtained from Experiment 7.

The ratings were overall lower than those found in Experiment 7, suggesting that participants did not, on average, believe to have been exposed to the critical verbal metaphors too often before. The mean familiarity rating was 28.5 and the median 26.17. The minimum score was 10.6 and the maximum was 53.7. The mean familiarity values for every item can be seen in Table 5.4.

### 5.4.3 Experiment 9: conventionality of verbal metaphors

Experiment 9 intended to isolate participants' subjective frequency of exposure to the metaphoric meaning of the verbs used in Experiment 2. To do this, the study consisted of two parts: In the first one, a group of participants gave their interpretation of the meaning of each verb as it appeared in is metaphoric meaning embedded in the corresponding critical sentence. I then analyzed the answers that participants provided and determined the most common one for each item. In the second part of Experiment 9 a different group of participants rated the frequency in which the verbs, deprived of context, were believed to be used with the metaphoric meaning obtained from the results of the first part of the experiment for every item.

**Participants**

Two separate groups of participants were used for each part of the study. The first group consisted of 50 monolingual German native speakers (ages 18-31), who were recruited using the "clickworker" recruiting platform. None of them participated in Experiments 7 or 8. The second group was made up of 590 monolingual German native speakers (ages 18-31).

The stark difference in number of participants between part 2 of Experiment 9 and all other ratings studies in this chapter was a consequence of the materials used for this set of experiments (i.e. the 36 verbal metaphors presented in chapter three). All of the 36 verbal metaphors were conceptually related: They all had similar verbs of physical containment expressing abstract difficulty. As a consequence of this, the participants in part 1 of Experiment 9 all gave very similar interpretations of the metaphors, resulting in almost half of the items having the same interpretation (see the Results section below). I therefore decided, after conducting part 1, to run part 2 of Experiment 9 as a single-item study: Participants saw only a single one of the 36 critical items, together with their most common interpretation obtained in part 1. As a consequence, the number of participants had to be increased from 50 in order to get multiple ratings for every critical item. The total number of participants of part 2 (590) was determined by monetary constraints.

All participants gave their informed consent prior to taking part in the study and received 3 Euros (participants in part 1) or 10 cents (participants in part 2) for their participation. This study was covered by the ethics vote of the psycholinguistics lab of the Humboldt-Universität zu Berlin.

**Materials and design**

In part one, participants were asked to interpret the metaphoric sentences, with the only constraint being brevity: They were asked to interpret the metaphors in as few words as possible. Every participants saw the 36 experimental sentences and 4 filler sentences intended as catch trials. These sentences were completely nonsensical. Participants were instructed to mark nonsensical sentences as such by writing "I do not understand this sentence" (or similar). Participants who did not identify at least 3/4 of nonsensical sentences as such were excluded from further analysis.
Given that all 36 items were conceptually related (all of the sentences conceptualized difficulty using a verb of spatial containment, see chapter 3 for details), the metaphoric interpretation of the items overlapped: For more than half of the items, the most common response in part one of the study was identical (see Table 5.2 below for the most common response for each item). Therefore, in order to avoid repetition effects, part two was conducted as a single item study: Participants saw just one critical item together with one nonsensical filler item. Participants who rated the single nonsensical filler with a score higher than 20 were discarded from the analysis.

In part two, Participants saw the de-contextualized target verb together with the most common metaphoric interpretation for that verb obtained in part one of the experiment. They were instructed to rate how often they believed to have heard or read this specific verb with the given metaphoric interpretation on a scale from 1 (never) to 100 (very often).

**Procedure**

In part one, participants read the sentence and saw a dialogue box underneath, where they could write their interpretation of the metaphor. After they had entered the interpretation, they could click on the "continue" button under the dialogue box. The entire experiment lasted around 15 minutes. In part two, participants saw the target verb written in bold with the intended metaphoric interpretation underneath. They were asked to rate how often they had encountered this verb with this meaning.

### Analysis and results

Prior to analyzing the results of part 1, the data from 4 participants was removed given their answers to the filler trials. After this, the answers of all participants were manually counted to determine the most common response. When participants answered with the same verb but in different tenses, or using the same verb but in a different syntactic construction, this was counted as being the same answer (e.g. *begrenzen begrenzt* and *nach oben begrenzt* were counted as belonging to the same class). The same was done for synonymous answers (e.g. *reduziert* und *verkleinert* were counted as belonging to the same class). For part 2, the data from 249 participants was removed from the analysis given their rating of the filler item. This reduced the total number of participants to 341.

The results of part 1 showed that participants chose the word *eingeschränkt* as the most appropriate interpretation of the metaphoric meaning for 16 out of the 36 items. This was followed by *verringert*, which was chosen 6 times, and *begrenzt*, which was chosen 4 times. The most common response for every item, which was used for part 2 of the experiment, can be seen in Table 5.2. In Table 5.2, "Critical_Word" refers to the target item that participants were asked to define, "Interpretation" is the most common response taht participants gave for each item.

The results of part 2 showed that on average, participants did not consider the metaphoric meaning of the verbs to be very conventional. The mean score was 37.771 and the median score was 31.1. The maximum rating was 61.89 and the minimum was 9.87. The final conventionality mean score for every item can be seen in Table 5.4.

| Item | Intepretation | Critical_word |
|---|---|---|
| 1 | begrenzt | gedeckelt |
| 2 | gedämpft | gedeckelt |
| 3 | begrenzt | ummauert |
| 4 | eingeschränkt | ummauert |
| 5 | eingeschränkt | gefesselt |
| 6 | eingeschränkt | umrandet |
| 7 | eingeschränkt | umrandet |
| 8 | eingeschränkt | umwickelt |
| 9 | verringert | umgrenzt |
| 10 | eingeschränkt | umgrenzt |
| 11 | eingeschränkt | umgittert |
| 12 | eingeschränkt | umgittert |
| 13 | eingeschränkt | umzäunt |
| 14 | eingeschränkt | umzäunt |
| 15 | eingeschränkt | beengt |
| 16 | verringert | beengt |
| 17 | verringert | beengt |
| 18 | eingeschränkt | eingeschnürt |
| 19 | begrenzt | eingekapselt |
| 20 | eingeschränkt | eingekapselt |
| 21 | eingeschränkt | einbetoniert |
| 22 | entzogen | einbetoniert |
| 23 | abgelehnt | eingeklemmt |
| 24 | eingeschränkt | eingeklemmt |
| 25 | verringert | festgekettet |
| 26 | eingeschränkt | festgekettet |
| 27 | gedämpft | festgebunden |
| 28 | eingeschränkt | festgebunden |
| 29 | begrenzt | abgesteckt |
| 30 | verringert | abgesteckt |
| 31 | eingeschränkt | eingedämmt |
| 32 | verringert | eingedämmt |
| 33 | eingeschränkt | abgeriegelt |
| 34 | verringert | abgeriegelt |
| 35 | beendet | abgesperrt |
| 36 | gedämpft | abgesperrt |

**Table 5.2:** Most common answer for every item in part 1 of Experiment 9.

### 5.4.4 Experiment 10: effects of aptness, familiarity and conventionality on processing verbal metaphors

#### Participants, materials, design, and procedure

Experiment 10 was a reanalysis of the data collected in Experiment 2. Specifically, of total sentence reading times and of the reading times in the VERB region, which were originally discussed in chapter three and are summarized in Table 3.7.

#### Predictions

The predictions for Experiment 10 were derived from the theoretical discussion on the role of aptness, conventionality and familiarity during metaphor processing: Whereas Category Inclusion views hold that aptness is the main factor that modulates processing mode (e.g. Glucksberg, 2008), Implicit Comparison views reserve this role for either conventionality (Bowdle & Gentner, 2005) or familiarity (Thibodeau & Durgin, 2011). This means that if the Category Inclusion view holds, we should find that aptness is a better fit to the reading times data in the VERB region of Experiment 2. If, on the other hand, the Implicit Comparison View holds, conventionality or familiarity should be a better fit to the data.

In terms of the point in time in which these effects should appear, we would expect to find them in total sentence reading times, as has been evidenced in the previous literature. Additionally, it should be possible to find effects on reading times of the metaphoric vehicle (the VERB region, see sentence (3) above), since at this point participants have read the entire metaphor. If the mediating factors have a rapid effect, we should be able to find that the mediating factors modulate first-pass reading times and/or regression path duration. If, on the other hand, these effects are somewhat delayed, we should find them in total reading times only.

#### Analysis

Since the goal of Experiment 10 was to disentangle the effects of aptness, conventionality and familiarity of reading times of verbal metaphors, it was important to try to minimize the correlation between these three variables for the set of materials.

I attempted to do this by (1) collecting ratings for each measure from different sets of participants, (2) increasing the number of data points compared to the original norming task, and (3) including a continuous slider scale to obtain more fine-grained results. However, this approach can only do so much to reduce correlation values, given that the materials were not designed with this purpose in mind. It can therefore be the case that some of the critical items happen to have similar values in the three categories. This was what was observed, as can be seen in Table 5.3 below.

**Table 5.3:** Correlation values of the ratings collected in Experiments 7-9

| *Correlation of ratings* | **Aptness** | **Familiarity** | **Conventionality** |
|---|---|---|---|
| **Aptness** | 1 | 0.8062579 | 0.5456825 |
| **Familiarity** | 0.8062579 | 1 | 0.6366677 |
| **Conventionality** | 0.5456825 | 0.6366677 | 1 |

The correlation values shown in Table 5.3 make it likely that fitting a model including all three of the variables at the same time will result in multicollinearity. Multicollinearity can lead to an inflated variance, distorting the individual contribution of each independent variable, which in turn makes it hard to interpret the results (see York, 2012; or McElreath, 2020, pg. 141-144). Because of this, four separate models were fitted to the data in order to explore the individual contribution of each variable. The first model - which served as a base level - did not include any of the three variables. It had the following structure, provided in R syntax below:

model 1: *Transformed reading times ~ Prime Type + Trial Order + (1 + Prime Type |item) + (1 + Prime Type|subject)*

Note that this model differs somewhat from the final model shown in Table 3.7. The present model removed the factor "Containment" given that this factor did not have a significant effect on the reading times and did not significantly improve the model fit. Including the factor "Containment" in the fixed and random effects to all the models in Experiment 10 does not change the results nor their interpretation in any way. The random effects structure was kept identical across all models.

Models 2-4, which included aptness, conventionality and familiarity, respectively were identical to the base model shown above with the exception of adding each one of the variables of interest individually. This means that models 2-4 only tested a main effect of the mediating factors and not their interaction with Prime Type or with Trial Order.

Models 2-4 were individually compared to model 1 through a log-likelihood test. To assess the fit of each model and therefore the unique contribution of each of the mediating factors (besides testing for significance), I took into consideration the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) of each model. Both of these measures index the degree to which the fixed and random variables can account for the dependent measure: The lower the AIC and BIC, the better the fit of the model to the data. Additionally, I also compared each model's marginal $R^2$ value, which assesses the percentage of variance explained by a model's fixed effects only. By comparing the values of each model in these categories, we can have a better idea of which model was overall best, in addition to the criterion of significance testing. All of these analyses were exploratory in nature. For completeness purposes, a final regression model was fitted to the data including all three mediating factors simultaneously.

This model-fitting procedure was conducted on the first-pass reading times, regression-path duration and total reading times of the VERB region, as well as on total sentence reading times. Only the VERB region was selected since it was the only one were it would be theoretically possible to expect an effect of all three mediating factors, since aptness and familiarity ratings were collected on the entire metaphoric sentences, but conventionality ratings were collected only for the verb. Total sentence reading times were considered to test for the possibility that effects would only be visible after the entire sentence had been understood.

**Results**

The results of the analyses show that the model including aptness as a factor was the overall best fit to the data. This was found for both total sentence reading times and total reading times of the VERB region, but not for first-pass reading times or for regression path duration, where there were no significant effects of any of the three mediators.

For sentence reading times and total reading times of the VERB region, the effect of aptness was significant, while there was no significant effect of familiarity or conventionality. This was also the case if Containment was included in the model. Adding the mediating factors did not affect any of the original results reported in chapter three.

The effect of aptness shows that metaphors that were rated as more apt were read faster than those rated as less apt. The model that included aptness had the highest marginal $R^2$ value out of the three models as well as the lowest values for AIC and BIC. The model including all mediating factors simultaneously also shows this same pattern for both total sentence reading times and total reading times in the VERB region: Only aptness had a significant facilitation effect on reading times. This model is shown in Table 5.9.

Because of the problem of multicollinearity, it was necessary to assess the validity of the results of the model including all variables simultaneously, depicted in Table 5.9. One way of doing this is by measuring a model's Variance inflation Factor (VIF) (Mansfield & Helms, 1982), which measures the relationships between the variance of a model with multiple terms and that of the same model with just one term. As a general rule of thumb, it is commonly stated that if the VIF is below 5, multicollinearity should not pose a problem for the interpretation of the regression coefficients (Neter et al., 1996). The VIF values for aptness, familiarity and conventionality for the model shown in Table 5.9 were all below 5, suggesting that the regression coefficients are reliable.

No significant effects were found on first-pass reading times or on regression-path duration of the VERB region.

These results can be visualized in Figures 5.3 and 5.4, and the output of all models is shown in Tables 5.5, 5.6, 5.7 and 5.8.

| item | familiarity | aptness | conventionality |
|---|---|---|---|
| 1 | 52.45 | 59.12 | 32.45 |
| 2 | 41.76 | 54.77 | 29.69 |
| 3 | 29.94 | 44.73 | 31.33 |
| 4 | 39.45 | 43.58 | 35.33 |
| 5 | 36.79 | 44.90 | 48.88 |
| 6 | 13.70 | 39.40 | 34.55 |
| 7 | 19.33 | 33.20 | 34.55 |
| 8 | 9.42 | 33.38 | 14.90 |
| 9 | 16.82 | 43.30 | 9.88 |
| 10 | 17.85 | 43.00 | 36.17 |
| 11 | 11.42 | 40.58 | 20.15 |
| 12 | 12.09 | 46.77 | 20.15 |
| 13 | 17.91 | 40.25 | 26.27 |
| 14 | 22.09 | 54.50 | 26.27 |
| 15 | 46.85 | 56.85 | 61.42 |
| 16 | 21.82 | 48.95 | 30.65 |
| 17 | 42.33 | 61.58 | 46.50 |
| 18 | 44.58 | 64.08 | 46.50 |
| 19 | 15.45 | 36.98 | 49.36 |
| 20 | 23.06 | 49.15 | 36.54 |
| 21 | 20.30 | 44.05 | 22.82 |
| 22 | 27.64 | 48.33 | 13.67 |
| 23 | 13.94 | 40.62 | 13.45 |
| 24 | 16.97 | 42.85 | 37.88 |
| 25 | 29.09 | 42.88 | 15.08 |
| 26 | 27.79 | 46.73 | 40.00 |
| 27 | 24.64 | 45.62 | 22.33 |
| 28 | 38.79 | 56.15 | 44.90 |
| 29 | 36.06 | 46.55 | 61.89 |
| 30 | 25.52 | 42.00 | 18.12 |
| 31 | 51.91 | 53.60 | 39.64 |
| 32 | 54.67 | 68.72 | 60.33 |
| 33 | 41.00 | 53.50 | 52.50 |
| 34 | 39.61 | 43.65 | 31.75 |
| 35 | 11.61 | 37.20 | 15.62 |
| 36 | 12.82 | 35.08 | 14.64 |

**Table 5.4:** Ratings collected in Experiments 7-9

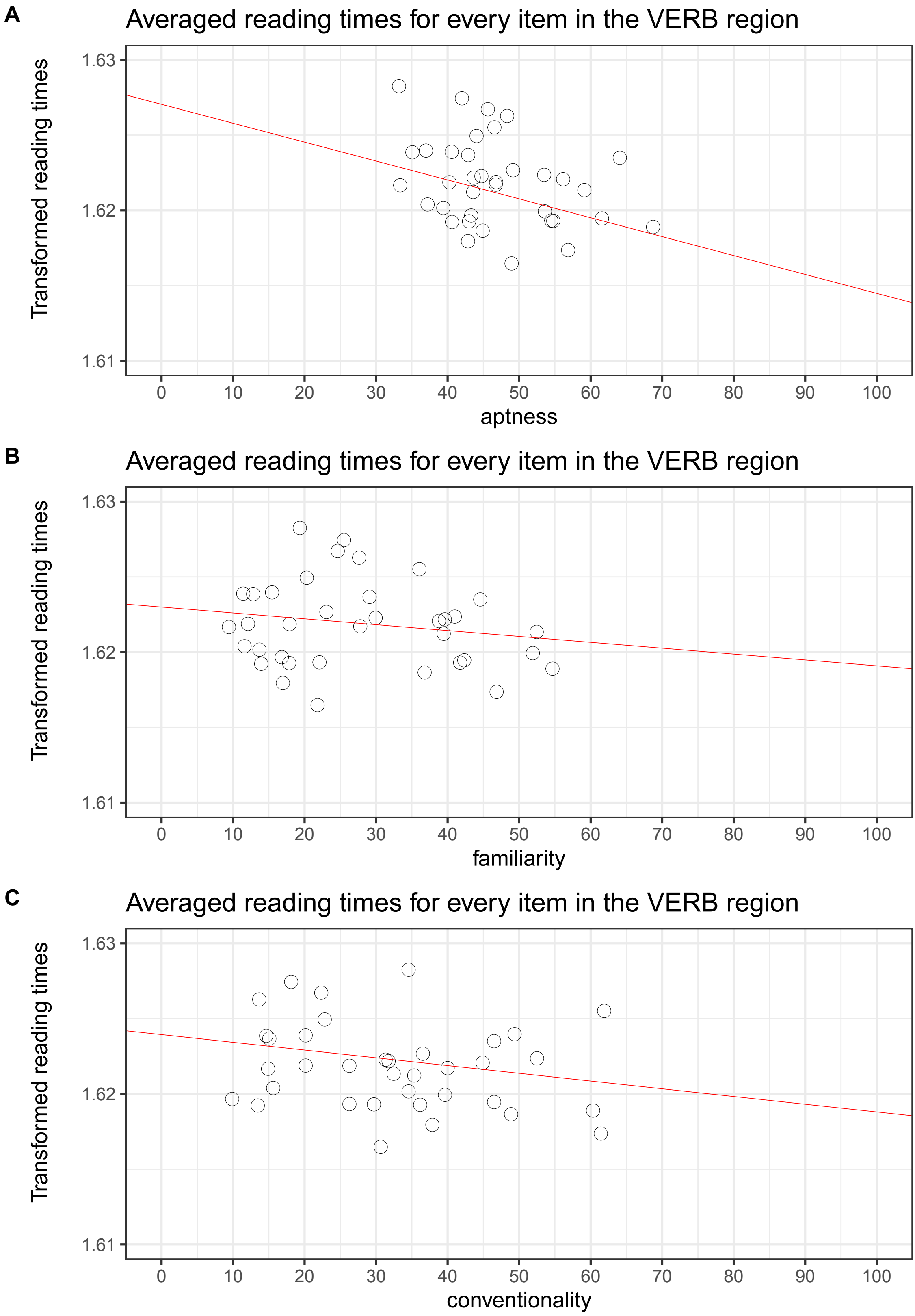


**Figure 5.3:** Mediating factors for total reading times of the verb region of Experiment 10

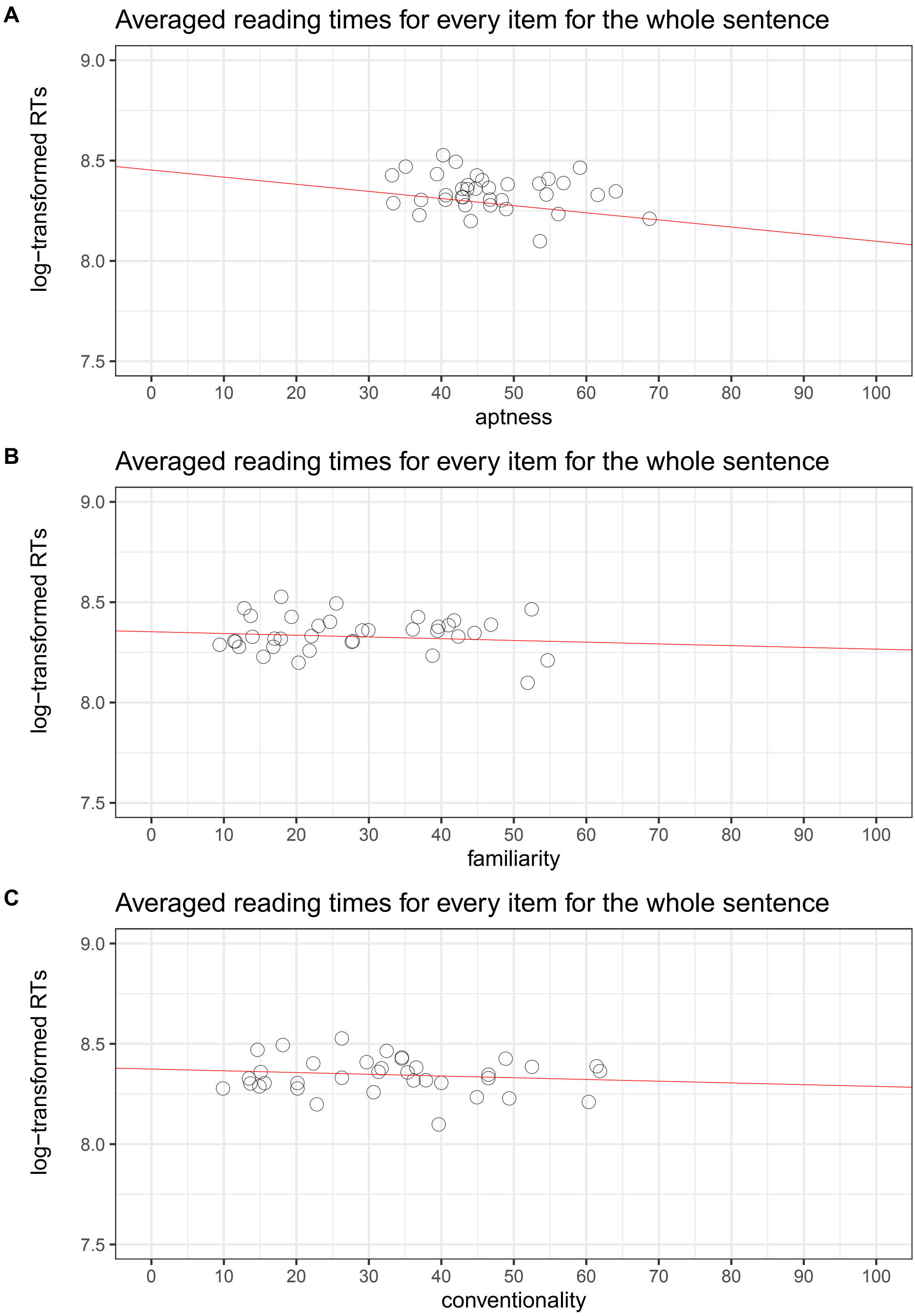


**Figure 5.4:** Mediating factors for reading times of the whole sentence of Experiment 10

**Table 5.5:** Mediating factors for first-pass times of Experiment 2, VERB region

| | *Dependent variable:* | | |
|---|---|---|---|
| | First-pass Reading Times | | |
| | (1) | (2) | (3) |
| Prime Type | −0.001<br>(−0.002, 0.0005)<br>t = −1.165 | −0.001<br>(−0.002, 0.0005)<br>t = −1.171 | −0.001<br>(−0.002, 0.001)<br>t = −1.161 |
| Aptness | 0.00002<br>(−0.0001, 0.0001)<br>t = 0.286 | | |
| Familiarity | | −0.00000<br>(−0.0001, 0.0001)<br>t = −0.110 | |
| Conventionality | | | 0.00002<br>(−0.0001, 0.0001)<br>t = 0.641 |
| Trial Order | −0.00004<br>(−0.0001, −0.00002)<br>t = −4.538*** | −0.00004<br>(−0.0001, −0.00002)<br>t = −4.549*** | −0.00004<br>(−0.0001, −0.00002)<br>t = −4.555*** |
| Intercept | 1.612<br>(1.606, 1.619)<br>t = 489.847*** | 1.613<br>(1.610, 1.617)<br>t = 943.229*** | 1.613<br>(1.609, 1.616)<br>t = 918.214*** |
| Marginal R-squared | 0.00985 | 0.00962 | 0.01026 |
| Observations | 1,566 | 1,566 | 1,566 |
| Log Likelihood | 4,679.936 | 4,679.902 | 4,680.093 |
| Akaike Inf. Crit. | −9,341.873 | −9,341.804 | −9,342.186 |
| Bayesian Inf. Crit. | −9,293.666 | −9,293.597 | −9,293.980 |

*Note:* *p<0.1; **p<0.05; ***p<0.01
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table 5.6:** Mediating factors for regression-path duration of Experiment 2, VERB region

| | *Dependent variable:* | | |
|---|---|---|---|
| | Regression Path Duration | | |
| | (1) | (2) | (3) |
| Prime Type | −0.002 | −0.002 | −0.002 |
| | (−0.004, −0.001) | (−0.004, −0.001) | (−0.004, −0.001) |
| | t = −3.964*** | t = −3.964*** | t = −3.964*** |
| Aptness | −0.00001 | | |
| | (−0.0001, 0.0001) | | |
| | t = −0.081 | | |
| Familiarity | | −0.00000 | |
| | | (−0.0001, 0.0001) | |
| | | t = −0.063 | |
| Conventionality | | | −0.00001 |
| | | | (−0.0001, 0.0001) |
| | | | t = −0.324 |
| Trial Order | −0.00004 | −0.00004 | −0.00004 |
| | (−0.0001, −0.00002) | (−0.0001, −0.00002) | (−0.0001, −0.00002) |
| | t = −4.522*** | t = −4.522*** | t = −4.521*** |
| Intercept | 1.616 | 1.616 | 1.616 |
| | (1.609, 1.623) | (1.613, 1.619) | (1.613, 1.620) |
| | t = 454.164*** | t = 917.532*** | t = 889.588*** |
| Marginal R-squared | 0.01689 | 0.01689 | 0.01709 |
| Observations | 1,566 | 1,566 | 1,566 |
| Log Likelihood | 4,629.192 | 4,629.191 | 4,629.242 |
| Akaike Inf. Crit. | −9,240.385 | −9,240.383 | −9,240.483 |
| Bayesian Inf. Crit. | −9,192.178 | −9,192.176 | −9,192.277 |

*Note:* *p<0.1; **p<0.05; ***p<0.01
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table 5.7:** Mediating factors for total reading times of Experiment 2, VERB region

| | *Dependent variable:* | | |
|---|---|---|---|
| | total Reading Times | | |
| | (1) | (2) | (3) |
| Prime Type | −0.003 | −0.003 | −0.003 |
| | (−0.004, −0.002) | (−0.004, −0.002) | (−0.004, −0.002) |
| | t = −4.566*** | t = −4.587*** | t = −4.591*** |
| Aptness | −0.0001 | | |
| | (−0.0002, −0.00002) | | |
| | t = −2.313** | | |
| Familiarity | | −0.00004 | |
| | | (−0.0001, 0.00003) | |
| | | t = −1.032 | |
| Conventionality | | | −0.0001 |
| | | | (−0.0001, 0.00002) |
| | | | t = −1.496 |
| Trial Order | −0.0001 | −0.0001 | −0.0001 |
| | (−0.0001, −0.0001) | (−0.0001, −0.0001) | (−0.0001, −0.0001) |
| | t = −9.537*** | t = −9.500*** | t = −9.483*** |
| Intercept | 1.633 | 1.629 | 1.629 |
| | (1.628, 1.639) | (1.625, 1.632) | (1.626, 1.632) |
| | t = 567.421*** | t = 993.223*** | t = 979.076*** |
| Marginal $R^2$ | 0.05191 | 0.04845 | 0.05022 |
| Observations | 1,566 | 1,566 | 1,566 |
| Log Likelihood | 4,608.710 | 4,606.728 | 4,607.223 |
| Akaike Inf. Crit. | −9,199.419 | −9,195.455 | −9,196.446 |
| Bayesian Inf. Crit. | −9,151.213 | −9,147.249 | −9,148.240 |

*Note:* *p<0.1; **p<0.05; ***p<0.01
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table 5.8:** Mediating factors for reading times of the whole sentence in Experiment 2

| | *Dependent variable:* | | |
|---|---|---|---|
| | Sentence Reading Times | | |
| | (1) | (2) | (3) |
| Prime Type | −0.099 | −0.099 | −0.099 |
| | (−0.121, −0.077) | (−0.121, −0.077) | (−0.121, −0.077) |
| | t = −8.715*** | t = −8.859*** | t = −8.866*** |
| Aptness | −0.004 | | |
| | (−0.007, −0.0005) | | |
| | t = −2.276** | | |
| Familiarity | | −0.001 | |
| | | (−0.003, 0.001) | |
| | | t = −0.819 | |
| Conventionality | | | −0.001 |
| | | | (−0.003, 0.001) |
| | | | t = −0.895 |
| Trial Order | −0.003 | −0.003 | −0.003 |
| | (−0.003, −0.003) | (−0.003, −0.003) | (−0.003, −0.003) |
| | t = −16.751*** | t = −16.707*** | t = −16.696*** |
| Intercept | 8.677 | 8.534 | 8.538 |
| | (8.519, 8.835) | (8.446, 8.623) | (8.447, 8.629) |
| | t = 107.418*** | t = 188.658*** | t = 183.833*** |
| Marginal R-squared | 0.09879 | 0.09345 | 0.09459 |
| Observations | 1,566 | 1,566 | 1,566 |
| Log Likelihood | 87.646 | 85.630 | 85.701 |
| Akaike Inf. Crit. | −153.292 | −149.261 | −149.402 |
| Bayesian Inf. Crit. | −94.373 | −90.341 | −90.483 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) | | |

**Table 5.9:** Mediating factors for both VERB region and sentence.

| | *Dependent variable:* | |
|---|---|---|
| | VERB Total RTs | Sentence RTs |
| | (1) | (2) |
| Prime Type | −0.003<br>(−0.004, −0.002)<br>t = −4.514*** | −0.099<br>(−0.121, −0.076)<br>t = −8.745*** |
| Aptness | −0.0002<br>(−0.0004, −0.00005)<br>t = −2.498** | −0.007<br>(−0.012, −0.002)<br>t = −2.672*** |
| Familiarity | 0.0001<br>(−0.00002, 0.0002)<br>t = 1.640 | 0.003<br>(−0.001, 0.006)<br>t = 1.644 |
| Conventionality | −0.00003<br>(−0.0001, 0.00005)<br>t = −0.813 | −0.0004<br>(−0.003, 0.002)<br>t = −0.347 |
| Trial Order | −0.0001<br>(−0.0001, −0.0001)<br>t = −9.495*** | −0.003<br>(−0.003, −0.003)<br>t = −16.755*** |
| Intercept | 1.636<br>(1.630, 1.643)<br>t = 489.505*** | 8.762<br>(8.579, 8.946)<br>t = 93.597*** |
| Observations | 1,566 | 1,566 |
| Log Likelihood | 4,610.018 | 88.982 |
| Akaike Inf. Crit. | −9,198.036 | −151.964 |
| Bayesian Inf. Crit. | −9,139.116 | −82.332 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01<br>values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) | |

### 5.4.5 Discussion

The results of Experiments 7-10 suggest that aptness modulates reading times of verbal metaphors better than conventionality and familiarity: After controlling for the effect of Prime Type and Trial Order, the aptness ratings collected in Experiment 7 showed that more apt metaphors were read significantly faster than less apt metaphors. This was the case for total sentence reading times, in line with the findings of Jones and Estes (2006), and for total reading times of the VERB region. The models including aptness also had an overall better fit than the models including conventionality and familiarity, who did not show any significant improvement on model fit compared to the base model. This suggests that aptness might be the best mediating factor during processing of verbal metaphors, and that this effect might only appear in secondary stages of metaphor comprehension.

At first glance, another interpretation seems possible: It could be the case that the differences in effects have to do with the differences in mean scores for aptness (mean value: 46.85), familiarity (mean value: 28.5) and conventionality (mean value: 37.771): Maybe aptness was found to modulate reading times because the items had an overall higher rating for aptness than for familiarity or conventionality. If another set of items had been used that had higher scores for familiarity and conventionality, it could have been possible to find a different pattern of effects. In other words, it could be that effects of a factor are only found when that factor has higher ratings vs. when the factors has lower ratings. However, one strong argument against this interpretation is the fact that both familiarity (min value: 10.6, max value: 53.7) and conventionality (min value: 9.8, max value:61.8) had a wider range than aptness did (min value: 33.20, max value: 68.72), and the ranges overlapped, particularly in the higher values. So if more conventional and more familiar metaphor would have shown an effect of conventionality and/or familiarity, the current set of materials would have been able to detect this.

One potential problem with the current results is the fact that there were considerable differences in the ratings collected for conventionality, on the one hand, and aptness and familiarity, on the other. Whereas the final ratings for conventionality came from an average of 9.5 individual ratings for every critical item, the ratings for aptness and familiarity came from an average of over 1200 individual ratings for every item, since for aptness and familiarity, every participant saw the full set of 36 items. This means that it is possible that the estimates for conventionality for every item are less accurate than the estimates for aptness and familiarity. To be sure that these estimates for conventionality are reliable, it would be necessary to collect ratings from a larger number of participants.

## 5.5 The Effect of Mediating Factors on Processing Spoken Verb-Object Metaphors

The following Experiments focus on the second type of non-nominal metaphors investigated in this dissertation: Verb-Object metaphors. The procedure here was identical to that of Experiments 7-10: I first collected ratings on the three dimensions of relevance (aptness, familiarity and conventionality) and then proceeded to fit the new values to the existing data collected in Experiment 5. Importantly, only a subset of the data was use for the current purposes: the metaphoric conditions of Experiment 5. This is due to the fact that the ratings collected collected on the three dimensions referred to the metaphorical interpretation of the sentences exclusively. Table 5.10 below presents an example of a context and Table 5.11 an example of a critical utterance for Experiment 5.

**Table 5.10:** Example of a linguistic context in the metphoric condition of Experiment 5

| Condition | Item |
|---|---|
| Metaphoric Context | Sebastian liebt eine wunderschöne Katze. Er hat sie in einem Tierheim adoptiert und seitdem sind sie unzertrennlich. Die Katze ist verwöhnt und launisch, und kann sehr wählerisch sein. Deswegen würde Sebastian alles für sie tun, wenn sie Hunger hat. Er wird sich immer um sie kümmern wollen. |

| **Condition** | **Item** |
|---|---|
| *English translation* | *"Sebastian loves a beautiful cat. He adopted her in a shelter and they have since been inseparable. The cat is spoiled and moody and can be very fussy. That's why Sebastian would do anything for her when she's hungry. He will always want to take care of her."* |

**Table 5.11:** Example of the metaphoric conditions of Experiment 5

| **Condition** | **Item** |
|---|---|
| Early verb position, metaphoric context | Sebastian füttert$_{\text{topic}}$ eine Prinzessin$_{\text{vehicle}}$, und wird unablässig der Katze beistehen. |
| *English Translation* | *"Sebastian will a princess feed and will relentlessly the cat support"* |
| Late verb position, metaphoric context | Sebastian wird eine Prinzessin$_{\text{vehicle}}$ füttern$_{\text{topic}}$, und wird unablässig der Katze beistehen. |
| *English Translation* | *"Sebastian will a princess feed and will relentlessly the cat support"* |

As described in chapter four, the verb-object metaphors of Experiment 5 were normed for aptness in order to make sure that the metaphors could be sufficiently understood. Based on this, two metaphors were discarded, leaving the total of critical items at 36. Additionally, ratings for familiarity were collected from the same group of participants, who were asked to state how often they had heard these metaphors before (if at all) on an ordinal scale from 1-7. Both of these ratings were used to compute a mean aptness and familiarity score for every item.

An initial exploratory analysis showed that aptness and familiarity might modulate processing: Metaphors that were rated more highly showed a higher log-gaze probability than the metaphors with lower aptness scores. This tentative finding was visible in the VEHICLE, VERB and UND regions.

However, these original ratings suffered from the same problems as the original ratings collected for Experiment 1: They used an ordinal measurement scale and were collected from only 20 participants. A further problem with this original norming study is the fact that aptness and conventionality were collected from the same sample of participants. As Jones and Estes (2006) argue, this can lead to the ratings being highly correlated: Participants tend to rate the metaphors that they consider to be more apt as more familiar and vice versa. This was in fact the case with these ratings, which had a correlation value of 0.826 for familiarity and aptness.

Given the theoretical importance of teasing the effects of aptness, familiarity and conventionality apart, it was paramount to obtain better quality data that could help do so. The collection of new ratings is described in Experiments 11-13. Their individual contribution to the processing of verbal metaphors is explored in Experiment 14. The ratings in Experiments 11-13 were all collected using the target items in the early-verb metaphoric condition only.

### 5.5.1 Experiment 11: aptness of verb-object metaphors

Experiment 11 is analogous to Experiment 7: it was meant to assess the perceived quality of the verb-object metaphoric expressions when embedded in their appropriate context. The Experiment was programmed using the IBEX software (Drummond, 2013) and run entirely online.

#### Participants

As with the previous Experiments in this chapter, 50 monolingual native speakers of German (ages 18-31) who did not participate in any other study in this chapter were recruited using the "clickworker" platform. All participants gave their informed consent prior to taking part in the study and received 2 Euros for their participation. This study was covered by the ethics vote of the psycholinguistics lab of the Humboldt-Universität zu Berlin.

#### Materials and design

The 36 critical items used in Experiment 5 served as the materials for Experiments 11-14. As with the previous studies in this chapter, participants saw 4 ungrammatical, completely non-sensical sentences which were used as attention controls: Only the data of participants who rated at least 3/4 of these sentences as incomprehensible

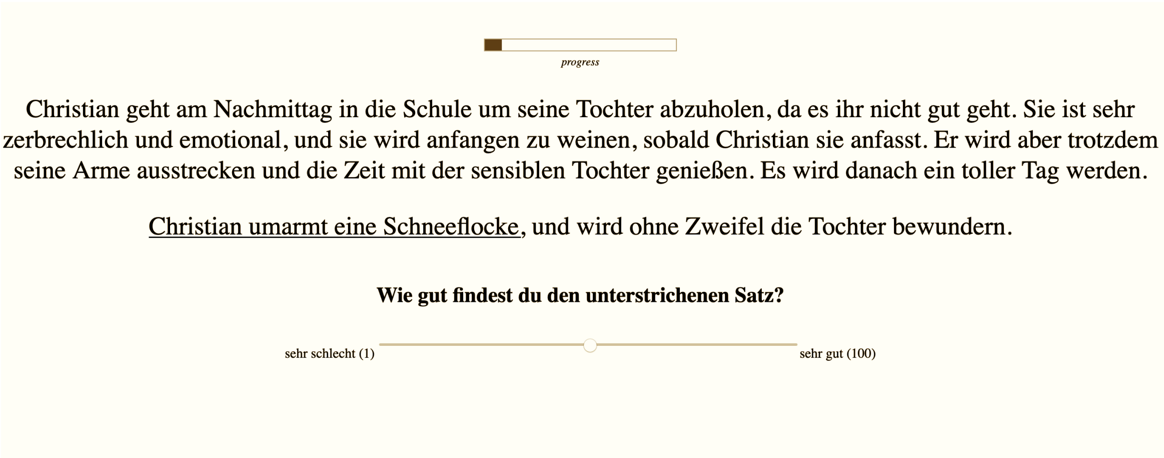


**Figure 5.5:** Example of a trial in Experiment 11

(i.e. with a value lower than 20 on the provided scale) were analyzed. The presentation of critical and filler items was randomized.

In the study, participants were asked to rate how understandable they thought every metaphor was given the provided context on a continuous, sliding scale from 1 (meaning totally incomprehensible) to 100 (meaning totally comprehensible). The metaphor was always underlined. Participants were not given an explicit definition of a metaphor. This was done in order to avoid biasing their answers as much as possible. An example critical trial can be seen in Figure 5.5.

### Procedure

On each trial, participants first read the 4-sentence context biasing the critical utterance towards a metaphoric interpretation. They then pressed the SPACE bar in order to read the metaphoric sentence. They pressed the SPACE bar again to reveal the slider scale. They changed the position of the slider button on the scale using their mouse, after which they clicked on the button “send”. The entire Experiment lasted around 15 minutes.

### Analysis and results

Prior to any analysis, the data from 13 participants was excluded based on their ratings of the filler items. This reduced the total number of participants to 37.

The data of these 37 participants was used to calculate the mean aptness value of each one of the 36 critical items. Only descriptive statistics were performed on the results of Experiment 11.

On average, participants rated the items to be somewhat comprehensible, with ratings being almost exactly half-way between the ends of the scale (mean value: 50.87, median value: 52.35). The lowest rating was 28.16 and the highest was 69.73. The mean aptness values for every item can be seen in Table 5.14.

### 5.5.2 Experiment 12: familiarity of verb-object metaphors

With this study, I collected data on perceived familiarity with verb-object metaphors.

#### Participants

50 monolingual native speakers of German (ages 18-31) who did not participate in any other study in this chapter were recruited via "clickworker". All participants gave their informed consent prior to taking part in the study and received 2 Euros for their participation. This study was approved by the ethics vote of the psycholinguistics lab of the Humboldt-Universität zu Berlin.

#### Materials, design, predictions and procedure

The materials, design, predictions and procedure were identical to those of Experiment 11 with the exception of the task: Participants were asked to rate the perceived frequency of the underlined verb-object metaphors on a scale from 1 (never before) to 100 (very often).

#### Analysis and results

Prior to the analysis, the data from 6 participants was excluded based on their ratings of the filler items. This reduced the total number of participants to 44. The data of these 44 participants was used to calculate the mean aptness value of each one of the 36 critical items. Only descriptive statistics were performed on the results of Experiment 12.

The results show that participants believed to be somewhat familiar with the set of metaphors. The mean rating was 45.67 and the median rating was 47.73. The spread was somewhat larger than in the previous experiment: The minimum rating was 17.86 and the maximum was 72.84. The mean familiarity values for every item can be seen in Table 5.14.

### 5.5.3 Experiment 13: conventionality of verb-object metaphors

As with Experiment 9, the goal of Experiment 13 was to obtain a measure of the metaphoric vehicle's de-contextualized metaphoric conventionality.

This Experiment was also divided into two parts: In the first one, a group of participants gave their interpretation of the meaning of each metaphor as it appeared embedded in the critical sentence (in the present tense) and with its corresponding 4-sentence context.

The most common answers per item were then determined and these were used as the basis of part two of the experiment, in which a different group of participants rated the frequency in which the target nouns, deprived of context, were believed to be used in the specific metaphoric meaning obtained in part 1.

#### Participants

Two separate groups of participants were used for each part of the study. The first group consisted of 50 monolingual German native speakers (ages 18-31), who were recruited using the "clickworker" recruiting platform. None of them participated in any of the other Experiments in this chapter. The second group also consisted of 50 monolingual German native speakers (ages 18-31).

All participants gave their informed consent prior to taking part in the study and received 3 Euros (participants in part 1) or 2 Euros (participants in part 2) for their participation. This study was approved by the ethics vote of the psycholinguistics lab of the Humboldt-Universität zu Berlin.

### Materials and design

In part one, participants were asked to interpret the metaphoric sentences in their provided context in as few words as possible. Every participants saw the 36 experimental sentences and 4 filler sentences intended as catch trials. These sentences were completely nonsensical. Participants were instructed to mark nonsensical sentences as such by writing "I do not understand this sentence" (or similar). Participants who did not identify at least 3/4 of nonsensical sentences as such were excluded from further analysis.

As opposed to Experiment 9, participants in part 2 of Experiment 13 also saw all 36 metaphoric items, given that these were not conceptually related in any particular way. They read the de-contextualized target noun (e.g. *Prinzessin*) together with the most common metaphoric interpretation for that noun obtained in part one of the experiment. They were instructed to rate how often they believed to have heard or read this specific noun with the given metaphoric interpretation on a scale from 1 (never) to 100 (very often).

### Procedure

The procedure was identical to that of Experiment 9.

### Analysis and results

No participant was removed from the analysis of part 1 given their responses to the filler trials. The answers of all participants were manually counted to determine the most common response. As in Experiment 9, participants' answers were grouped together with answers that were conceptually identical but superficially different (e.g. *langsame Spieler langsames Team* and *die Leute waren langsam* were counted as variations of the same response, namely *langsam*). Synonymous words were also grouped together (e.g. *gefühlslos* und *emotionslos* were counted as being the same response). For part 2, the data from 8 participants was removed from the analysis given their ratings of the filler items. This reduced the total number of participants to 42.

The results of part 1 showed that participants tended to provide more lengthy responses than those found in part 1 of Experiment 9, often providing more than one identifying property for the metaphor. Because of this, two stated characteristics were used as the final interpretation whenever more than half of the participants provided more than one identifying property. The final most common responses, which were chosen as the materials for part 2, are shown in Table 5.12: Here, "Critical_word" refers to the metaphoric vehicle that participants had to define. "Interpretation" refers to the average most common interpretation that participants gave to the metaphoric vehicles.

The results of part 2 showed that on average, participants considered the metaphoric meaning of the nouns to be somewhat conventional. The mean score was 54.9 and the median score was 55.06. The maximum rating was 96.26 and the minimum was 22.48.

### 5.5.4 Experiment 14: effects of aptness, familiarity and conventionality on processing verb-object metaphors

#### Participants, materials, design, and procedure

Experiment 14 was a reanalysis of the data collected in Experiment 5. Specifically, of the log-gaze ratios computed for the VEHICLE, VERB and UND time-windows (see chapter 4 for details on the original analyses performed on these regions). One important difference in the way the data was analyzed in Experiment 14 was the fact that a subset of the original data was taken containing only the metaphoric conditions. This was done because the ratings collected in Experiments 11-13 were based on the metaphoric conditions only.

#### Predictions

The predictions for Experiment 14, as with Experiment 10, were derived from the theoretical discussion on the role of aptness, conventionality and familiarity during metaphor processing: If the Category Inclusion view holds, it should be possible to find that aptness is a better fit to the data than familiarity and/or conventionality.

| item | Interpretation | Critical_word |
|---|---|---|
| 1 | rücksichtsloser und wütender Mensch | Bulldozer |
| 2 | wütender und gefährlicher Mensch | Drachen |
| 3 | emotionsloser Mensch | Roboter |
| 4 | Gruppe von Menschen, die langsam und träge sind | Altersheim |
| 5 | gütige, uneigennützige Person | Engel |
| 6 | hinterlistige, gierige Person | Hyäne |
| 7 | sehr alte und erfahrene Person | Dinosaurier |
| 8 | negative emotionale veränderung in einer Beziehung | Riss |
| 9 | aggressive und starke person | Gorilla |
| 10 | aggressive und gierige Person | Hai |
| 11 | sehr langsame person | Schildkröte |
| 12 | cholerische, unberechenbare person | Vulkan |
| 13 | charakterlich wandlungsfähiger Mensch | Chamäleon |
| 14 | starker, mutiger Mensch | Löwen |
| 15 | Ort, mit chaotischen Zuständen | Zirkus |
| 16 | jemand, der sich nicht zeigt und schwer aufzufinden ist | Gespenst |
| 17 | hartes, unverdauliches Essen | Stein |
| 18 | anmutige, elegante person | Schmetterling |
| 19 | eingängiger Song, der schnell an bekanntheit gewinnt | Virus |
| 20 | stark übergewichtiger mensch | Walross |
| 21 | zierliche, leicht verwundbare person | Schneeflocke |
| 22 | Schmächtiger und schwächlicher Mensch | Zahnstocher |
| 23 | Dummer Mensch der Glück hat | Esel |
| 24 | große, breitgebaute Person | Schrank |
| 25 | sehr alte Person mit veralteten Ansichten | Fossil |
| 26 | abweisende, gefühllose Person | Eisstatue |
| 27 | chaotischer, undurchsichtiger Ort | Dschungel |
| 28 | lindernde aber temporäre Lösung | Pflaster |
| 29 | athletische, sehr schnelle Person | Leoparden |
| 30 | hinterlistige, falsche Person | Schlange |
| 31 | großer, wütender mensch, der seine Meinung durrchsetzt | Panzer |
| 32 | verwöhnte und launische Person | Prinzessin |
| 33 | redselige Person | Papagei |
| 34 | agressive und gefährliche Person | Piranha |
| 35 | habgierige Mensch | Geier |
| 36 | eine schwierig zu erobernde Person, die nur als Statussymbol gilt. | Trophäe |

**Table 5.12:** Most common answer for every item in part 1 of Experiment 13

If, on the other hand, the Indirect comparison view holds, familiarity and/or conventionality should provide a better fit to the data than aptness.

A further prediction relates to the time in which effects should be visible. In Experiment 10, it was found that the effect of aptness was highly localized to the total reading times of the verb region, with no effect being found in first-pass or regression-path duration. In the Visual World Paradigm, participants do not have the option to revisit the linguistic input as they do when reading. Instead, they must continue to process the unfolding utterance and relate it to the visual context. If it is the case that effects of mediating factors should appear in later stages of metaphor comprehension, we would expect them to appear in the "und" region, as opposed to the "verb" or "vehicle" regions. If, on the other hand, these factors can appear earlier during sentence comprehension, we should see them in the vehicle and/or verb regions.

**Analysis**

As with Experiment 10, one of the goals of the newly collected ratings was to minimize the correlation values between variables. This was only partially successful: The correlation between aptness and familiarity was reduced from 0.8 to 0.66, but 0.66 is still a relatively high correlation value. All correlation values are presented in Table 5.13 below.

Because of the high correlation values, the analysis proceeded as in Experiment 10: Four separate models were fitted to the data in each of the three time windows (vehicle, verb, and "und") in order to explore the individual contribution of each variable. The baseline model did not include any of the three variables but had an identical random effects structure. This is provided in R syntax below:

> *log-gaze ratio of looks to metaphoric picture over literal picture ~ verb position + trial order + (1 +verb position| subject) + (1 +verb position| Item)*

Note that these models differ slightly from the final models used for these regions in chapter four. The present models were fitted only on the metaphoric conditions, seeing that the ratings were themselves only collected on the metaphoric levels of the factors "context bias" (they are, after all, measuring dimensions of metaphoricity, making them uninterpretable for the literal conditions). The factors "verb position" is still included in order to control for its effect during processing.

The models including aptness, conventionality and familiarity, were identical to the models presented above with the exception of adding each one of the variables of interest individually in each region. The models were compared to the corresponding base model through a log-likelihood test. the AIC, BIC and marginal $R^2$ of each model were then informally compared to one another to determine which of the three models had the best overall fit to the data. Finally, a full model including all three mediating factors simultaneously was fitted for every one of the three time windows.

**Table 5.13:** Correlation values of ratings collected in Experiments 11-13

| *Correlation of ratings* | **Aptness** | **Familiarity** | **Conventionality** |
|---|---|---|---|
| **Aptness** | 1 | 0.6585614 | 0.5245973 |
| **Familiarity** | 0.6585614 | 1 | 0.7643484 |
| **Conventionality** | 0.5245973 | 0.7643484 | 1 |

### Results

The results show that none of the three mediating factors significantly improved the model fit on the verb or on the vehicle region. For the vehicle region, the model including aptness showed a better fit to the data compared to the competing models, but not a better fit than the base model without any of the mediating variables. Similarly, in the verb region, the model including familiarity showed the best fit out of the three models with mediating variables, but it failed to be a better fit than the base model.

In the "und" region, only the model including aptness showed a significant improvement compared to the base model, with higher values of aptness associated to higher log-gaze probability ratios. This model also showed a lower BIC and AIC scores than the competing models, as well as a higher value of marginal $R^2$. This suggests that the more apt a metaphor, the more participants directed their gaze towards the metaphoric picture and less towards the literal picture. This finding also lends some support to the idea that item-specific properties, such as aptness, might have a delayed influence on processing, similarly to what was observed in Experiment 10.

All results of Experiment 14 can be visualized in Figures 5.6, 5.7 and 5.8. In the figures, the dotted blue line represents the point of no preference for viewing either literal or metaphoric picture. Positive values on the x-axis (i.e. above the blue line) represent a preference for the metaphoric picture and negative values (below the blue line) a preference for the literal picture. The red line shows the regression line. The output of all models is shown in Tables 5.15, 5.16, and 5.17. Table 5.18 shows the results of the models that included all variables simultaneously. As was the case in Experiment 10, the VIF values for this model were assessed in order to rule out a distortion of the coefficients because of multicollinearity. All VIF values in the model shown in Table 5.18 were below 5. These models thus suggest that only aptness has an effect on processing and that this effect is only visible in the “und” region.

One problem with the interpretation of these results comes from the fact that I examined only a subset of the original data. Whereas in Experiment 10 (which had 64 participants in total) the full set of 36 metaphors was analyzed, Experiment 14 consisted only of 18 items and 32 participants (since only the metaphoric conditions were considered for the analysis). It might therefore be the case that Experiment 14 was not sufficiently powered to detect effects of the mediating factors in the vehicle or verb regions, or to detect potential effects of conventionality and or familiarity in the “und” region. The current results should therefore be considered to be exploratory and need to be confirmed by a follow-up replication study.

### 5.5.5 Discussion

Experiments 11-13 collected ratings on the familiarity, aptness and conventionality of verb-object metaphors. In Experiment 14, these ratings were used to assess the individual contribution of each of these factors to the construction of metaphoric meaning. The findings showed that aptness was the only factor that showed an improved model fit compared to a base model. This effect was found only in the “und” region, which came immediately after the second element of the metaphor (the vehicle in the early-verb condition and the topic in the late-verb condition). The results suggest that, in the “und” region, participants preferred to inspect the metaphoric picture more than the literal picture if the metaphor was particularly apt. The less apt a metaphor, the more they preferred to look at the literal picture compared to the metaphoric picture. This is consistent with the idea that more apt metaphors should be more easily understood.

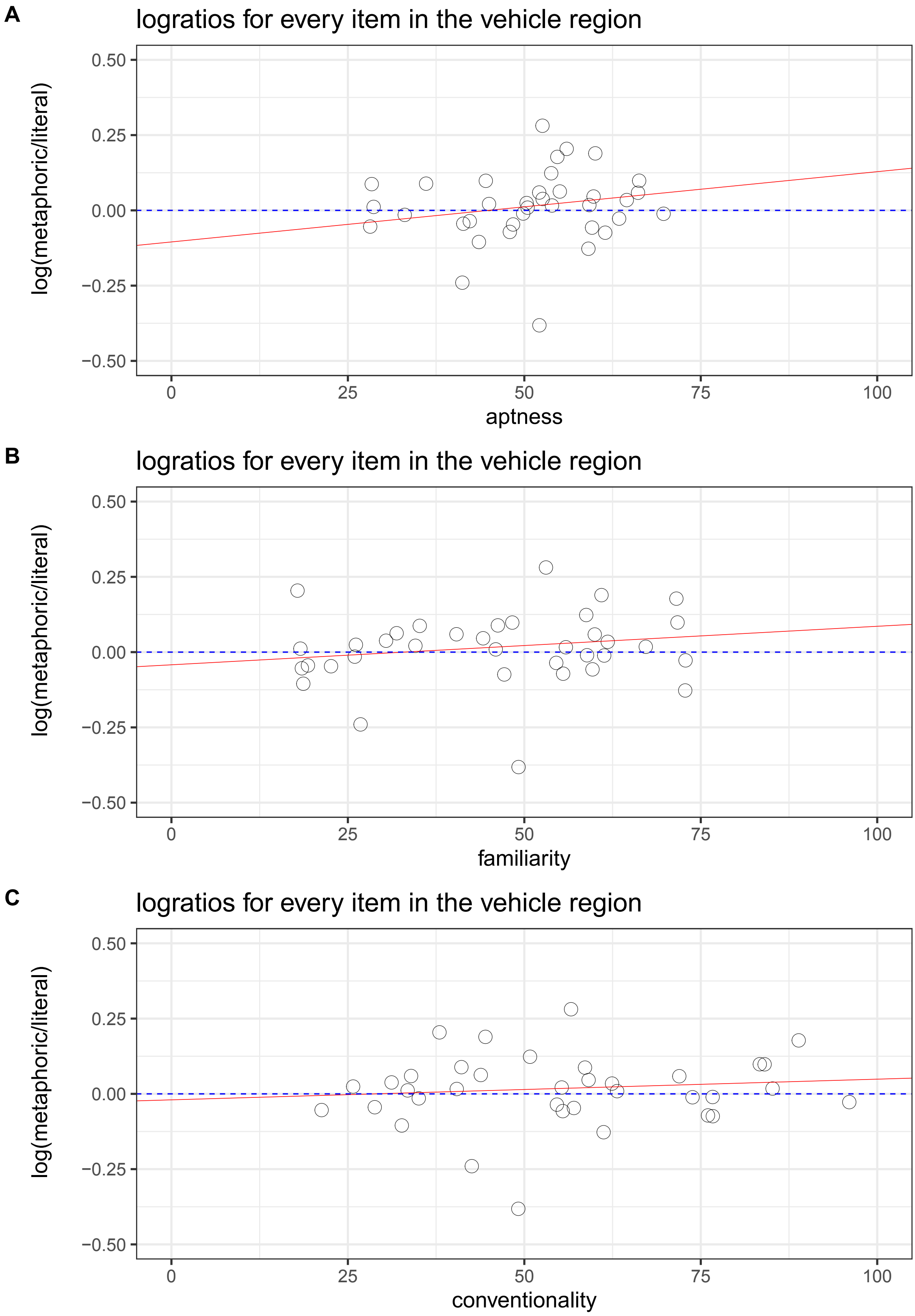


**Figure 5.6:** Mediating factors for log-gaze probability ratios in the vehicle region of Experiment 14

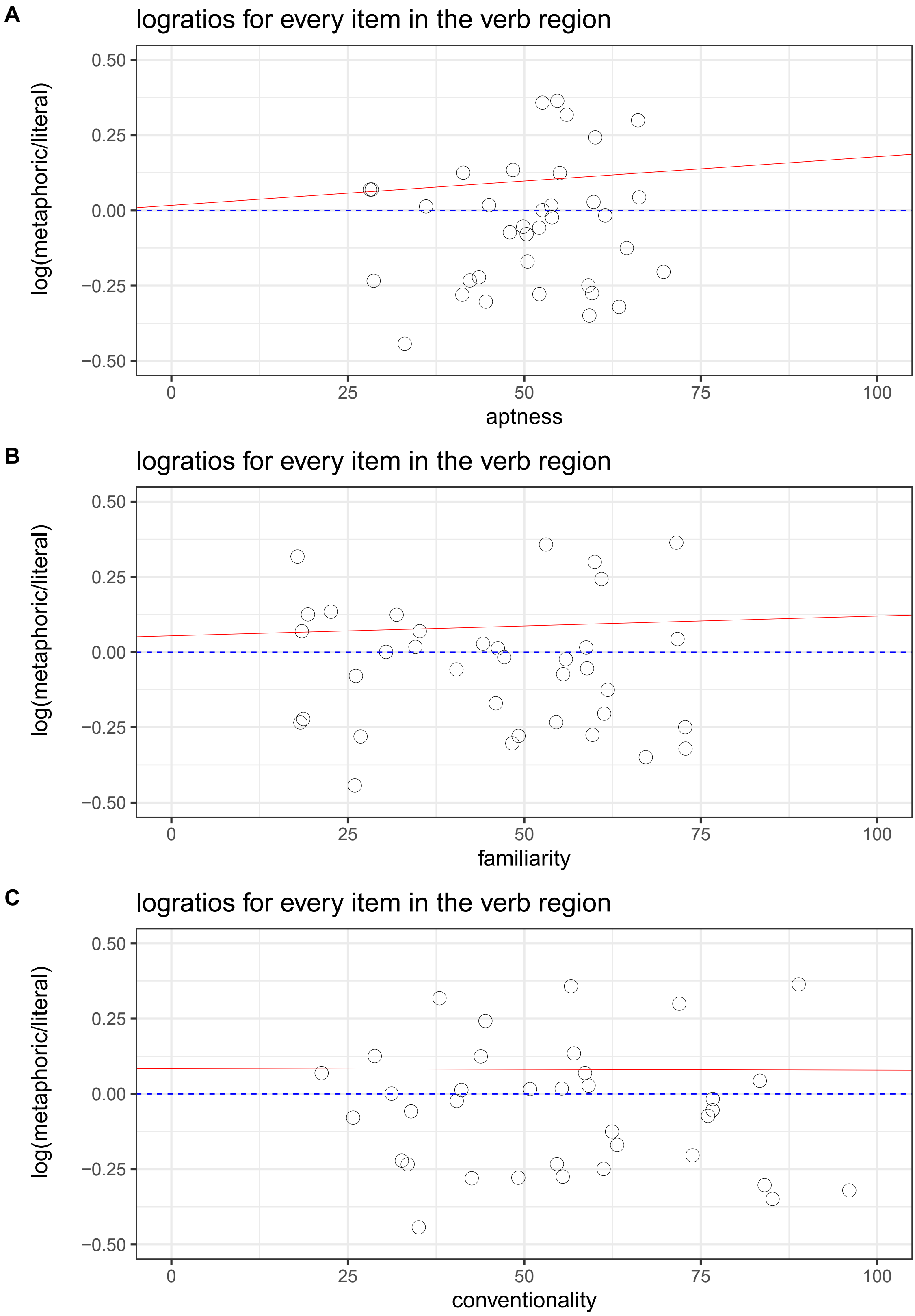


**Figure 5.7:** Mediating factors for log-gaze probability ratios in the verb region of Experiment 14

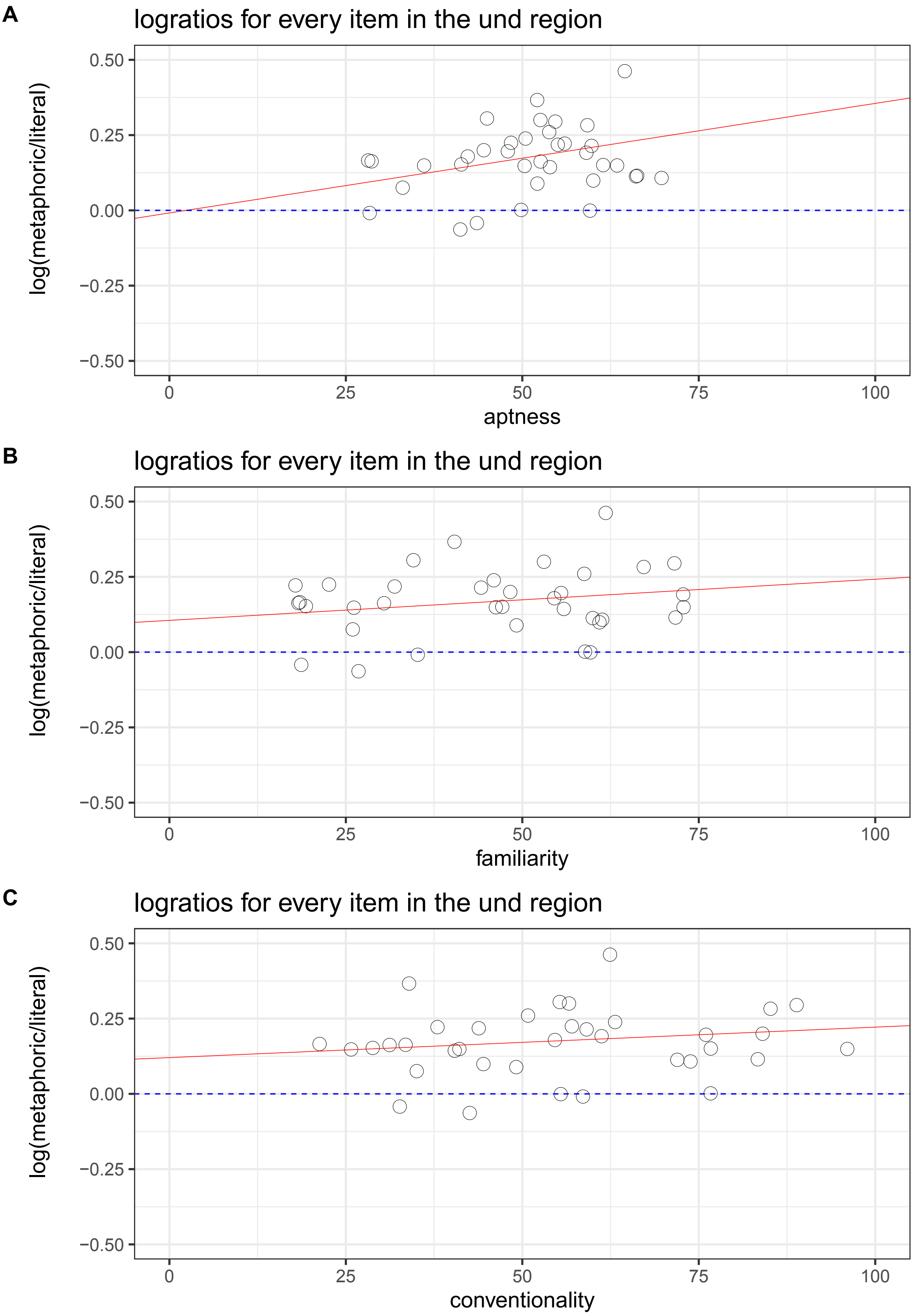


**Figure 5.8:** Mediating factors for log-gaze probability ratios in the 'und' region of Experiment 14

| Item | familiarity | aptness | conventionality |
|---:|---:|---:|---:|
| 1 | 26.14 | 50.32 | 25.74 |
| 2 | 72.80 | 59.08 | 61.23 |
| 3 | 55.50 | 47.95 | 76.02 |
| 4 | 26.80 | 41.22 | 42.55 |
| 5 | 67.20 | 59.22 | 85.17 |
| 6 | 31.91 | 55.03 | 43.83 |
| 7 | 55.89 | 53.92 | 40.43 |
| 8 | 44.16 | 59.81 | 59.11 |
| 9 | 54.52 | 42.27 | 54.62 |
| 10 | 60.93 | 60.05 | 44.49 |
| 11 | 47.16 | 61.46 | 76.70 |
| 12 | 25.98 | 33.05 | 35.04 |
| 13 | 34.57 | 45.00 | 55.30 |
| 14 | 59.98 | 66.11 | 71.96 |
| 15 | 61.82 | 64.51 | 62.43 |
| 16 | 59.66 | 59.59 | 55.45 |
| 17 | 35.18 | 28.38 | 58.60 |
| 18 | 40.39 | 52.11 | 33.96 |
| 19 | 18.48 | 28.16 | 21.28 |
| 20 | 58.89 | 49.84 | 76.68 |
| 21 | 30.41 | 52.59 | 31.19 |
| 22 | 49.18 | 52.14 | 49.15 |
| 23 | 46.25 | 36.08 | 41.09 |
| 24 | 71.55 | 54.68 | 88.87 |
| 25 | 58.77 | 53.81 | 50.81 |
| 26 | 18.68 | 43.57 | 32.64 |
| 27 | 45.95 | 50.46 | 63.15 |
| 28 | 22.61 | 48.41 | 57.02 |
| 29 | 17.86 | 56.00 | 37.98 |
| 30 | 72.84 | 63.43 | 96.04 |
| 31 | 18.27 | 28.65 | 33.47 |
| 32 | 71.70 | 66.27 | 83.36 |
| 33 | 53.07 | 52.57 | 56.62 |
| 34 | 19.34 | 41.35 | 28.81 |
| 35 | 48.30 | 44.54 | 84.04 |
| 36 | 61.32 | 69.73 | 73.83 |

**Table 5.14:** Ratings collected in Experiments 11-13

These results are less conclusive than those of Experiment 10, given the substantially reduced number of data points available for the analysis. Additionally, the two conditions of Experiment 14 changed the order of presentation of the metaphoric elements, meaning that when hearing the vehicle in one condition and

**Table 5.15:** Mediating factors for log-gaze probability ratios in the VEHICLE region of Experiment 5

| | *Dependent variable:* | | |
|---|---|---|---|
| | logratio | | |
| | (1) | (2) | (3) |
| Verb Constraint | 0.155 | 0.155 | 0.155 |
| | (0.038) | (0.038) | (0.038) |
| | t = 4.085*** | t = 4.084*** | t = 4.085*** |
| Aptness | 0.002 | | |
| | (0.002) | | |
| | t = 1.031 | | |
| Familiarity | | 0.001 | |
| | | (0.001) | |
| | | t = 0.951 | |
| Conventionality | | | 0.001 |
| | | | (0.001) |
| | | | t = 0.745 |
| Trial Order | 0.001 | 0.001 | 0.001 |
| | (0.0003) | (0.0003) | (0.0003) |
| | t = 4.927*** | t = 4.935*** | t = 4.941*** |
| Intercept | −0.102 | −0.059 | −0.053 |
| | (0.088) | (0.054) | (0.058) |
| | t = −1.158 | t = −1.103 | t = −0.926 |
| Marginal $R^2$ | 0.02013 | 0.01982 | 0.0194 |
| Observations | 41,280 | 41,280 | 41,280 |
| Log Likelihood | −32,878.880 | −32,878.960 | −32,879.120 |
| Akaike Inf. Crit. | 65,779.760 | 65,779.920 | 65,780.240 |
| Bayesian Inf. Crit. | 65,874.670 | 65,874.830 | 65,875.150 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 | | |
| | values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) | | |

**Table 5.16:** Mediating factors for log-gaze probability ratios in the VERB region of Experiment 5

| | *Dependent variable:* | | |
|---|---|---|---|
| | logratio | | |
| | (1) | (2) | (3) |
| Verb Constraint | 0.008 | 0.008 | 0.008 |
| | (−0.041, 0.056) | (−0.041, 0.056) | (−0.041, 0.056) |
| | t = 0.305 | t = 0.305 | t = 0.305 |
| Aptness | 0.0002 | | |
| | (−0.004, 0.004) | | |
| | t = 0.121 | | |
| Familiarity | | 0.0004 | |
| | | (−0.002, 0.003) | |
| | | t = 0.303 | |
| Conventionality | | | 0.0001 |
| | | | (−0.002, 0.002) |
| | | | t = 0.063 |
| Trial Order | 0.001 | 0.001 | 0.001 |
| | (0.001, 0.002) | (0.001, 0.002) | (0.001, 0.002) |
| | t = 5.151*** | t = 5.151*** | t = 5.152*** |
| Intercept | 0.037 | 0.033 | 0.046 |
| | (−0.165, 0.240) | (−0.084, 0.150) | (−0.082, 0.173) |
| | t = 0.363 | t = 0.551 | t = 0.704 |
| Marginal $R^2$ | 0.00096 | 0.00107 | 0.00093 |
| Observations | 41,760 | 41,760 | 41,760 |
| Log Likelihood | −32,538.690 | −32,538.650 | −32,538.690 |
| Akaike Inf. Crit. | 65,093.380 | 65,093.300 | 65,093.390 |
| Bayesian Inf. Crit. | 65,162.490 | 65,162.420 | 65,162.500 |

*Note:* *p<0.1; **p<0.05; ***p<0.01
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table 5.17:** Mediating factors for log-gaze probability ratios in the UND region of Experiment 5

| | *Dependent variable:* | | |
|---|---|---|---|
| | logratio | | |
| | (1) | (2) | (3) |
| Verb Constraint | −0.016 | −0.016 | −0.016 |
| | (−0.118, 0.086) | (−0.118, 0.086) | (−0.118, 0.086) |
| | t = −0.301 | t = −0.301 | t = −0.300 |
| Aptness | 0.004 | | |
| | (0.0004, 0.007) | | |
| | t = 2.199** | | |
| Familiarity | | 0.001 | |
| | | (−0.001, 0.003) | |
| | | t = 1.184 | |
| Conventionality | | | 0.001 |
| | | | (−0.001, 0.003) |
| | | | t = 1.074 |
| Trial Order | 0.001 | 0.001 | 0.001 |
| | (0.001, 0.002) | (0.001, 0.002) | (0.001, 0.002) |
| | t = 4.308*** | t = 4.326*** | t = 4.337*** |
| Intercept | −0.032 | 0.097 | 0.098 |
| | (−0.212, 0.149) | (−0.024, 0.219) | (−0.030, 0.225) |
| | t = −0.342 | t = 1.572 | t = 1.497 |
| Marginal R-squared | 0.00621 | 0.0024 | 0.00211 |
| Observations | 34,515 | 34,515 | 34,515 |
| Log Likelihood | −25,222.920 | −25,224.330 | −25,224.380 |
| Akaike Inf. Crit. | 50,467.830 | 50,466.670 | 50,470.770 |
| Bayesian Inf. Crit. | 50,560.770 | 50,542.710 | 50,563.710 |

*Note:* *p<0.1; **p<0.05; ***p<0.01
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table 5.18:** Mediating factors for log-gaze probability ratios for all regions simultaneously

| | *Dependent variable:* | | |
|---|---|---|---|
| | logratio | | |
| | (1) | (2) | (3) |
| Verb Constraint | 0.151 | 0.015 | 0.020 |
| | (0.077, 0.226) | (−0.082, 0.112) | (−0.077, 0.117) |
| | t = 3.966*** | t = 0.305 | t = 0.404 |
| Aptness | 0.0004 | −0.0002 | 0.002 |
| | (−0.004, 0.005) | (−0.005, 0.005) | (0.001, 0.003) |
| | t = 0.187 | t = −0.092 | t = 6.158*** |
| Familiarity | 0.0004 | 0.001 | 0.0001 |
| | (−0.003, 0.004) | (−0.003, 0.005) | (−0.0004, 0.001) |
| | t = 0.203 | t = 0.394 | t = 0.412 |
| Conventionality | 0.0003 | −0.0004 | 0.0002 |
| | (−0.003, 0.003) | (−0.004, 0.003) | (−0.0003, 0.001) |
| | t = 0.205 | t = −0.256 | t = 0.815 |
| Trial Order | 0.001 | 0.001 | 0.001 |
| | (0.001, 0.002) | (0.001, 0.002) | (0.001, 0.002) |
| | t = 4.233*** | t = 5.148*** | t = 3.895*** |
| Intercept | −0.069 | 0.048 | 0.022 |
| | (−0.249, 0.111) | (−0.161, 0.257) | (−0.008, 0.051) |
| | t = −0.752 | t = 0.451 | t = 1.456 |
| Observations | 41,280 | 41,760 | 34,515 |
| Log Likelihood | −33,879.150 | −32,538.610 | −26,894.950 |
| Akaike Inf. Crit. | 67,780.300 | 65,097.220 | 53,807.900 |
| Bayesian Inf. Crit. | 67,875.210 | 65,183.620 | 53,883.940 |

*Note:* *p<0.1; **p<0.05; ***p<0.01
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

the verb in the other they had not heard the entire metaphor yet. This might explain why results were only visible in the "und" region, which consistently appeared after the metaphor had been fully understood in both conditions.

However, there was a more balanced number of observations for the ratings of aptness, familiarity and conventionality, since the number of participants that rated each dimension was 37, 42, and 44 respectively. In Experiments 11-13, every participant saw the full set of items. This makes it easier to compare the three dimensions to one another.

## 5.6 Interim Conclusion

In this chapter, I investigated the effect of aptness, conventionality and familiarity on the processing of verbal metaphors and verb-object metaphors. These mediating factors play different roles in the theories that have been evaluated throughout this thesis: Whereas Implicit Comparison views attribute an important mediating role to conventionality and/or familiarity, Category Inclusion models view aptness as the true mediating factor.

Overall, the results present tentative support for the Category Inclusion view: For both verbal metaphors and verb-object metaphors, aptness was the only factor that significantly improved model fits. The effect were also found to be in a direction compatible with the theories: For verbal metaphors, more apt metaphors showed shorter reading times than less apt metaphors. For verb-object metaphors, more apt metaphors showed higher log-gaze ratios than less apt metaphors, which can be interpreted as higher apt metaphors displaying a more pronounced preference for visually inspecting the metaphoric picture vs. the literal picture compared to less apt metaphors.

Importantly, these effects were found to occur at somewhat later stages of metaphor processing for the investigated dependent measures: Experiment 10 found effects in total-reading times and sentence reading times (but not on first-pass or regression path duration), while Experiment 14 found effects in the "und" time window only. This suggests that Category Inclusion models need to be further specified to account for these late effects by postulating that aptness influences a later processing stage only.

It could, for example, be stated that aptness affects the ease in which the ad hoc category is ultimately constructed, but only after the relevant parameters for modulation have been determined from the metaphoric topic together with the utterance's context.

It is important to note, however, that these conclusions are only tentative. This is particularly the case for the interpretation of the results of Experiment 14, which might not have been sufficiently powered to answer the questions at hand, given that the analysis was only performed on half of the observations of the full data set. This is only speculative, and further analyses would be necessary to determine the minimum effect size of interest and the number of participants necessary to conduct an appropriately powered study investigating only the metaphoric conditions. This way, it should be possible to determine whether or not there are any effects of conventionality, aptness and familiarity and whether these effects appear any earlier than the "und" region.

One possible criticism to these results is that only effects for aptness were found because the ratings collected for aptness were a better estimate for the true value of aptness compared to the ratings collected for familiarity and conventionality. Since it might be difficult for participants to assess their perceived frequency of exposure, these ratings might have introduced more noise to the data than the aptness ratings did.

However, if this is the case, this problem might permeate the entire research area of mediating factors of metaphor comprehension, since the procedure followed here was identical to the one used in the two previous studies addressing the issue of theoretical repercussions of mediating effects (Bowdle & Gentner, 2005; Jones & Estes, 2006). Compared to these studies, the present experiments represent a methodological improvement by (1) testing more participants on average, (2) testing different types of metaphors and (3) measuring and keeping the mediating factors as continuous variables. It is therefore likely that the estimates for the mediating factors presented here represent an overall step in the right direction, even if the possibility remains that they might not be optimal. Future research could address this issue by testing the within-items reliability rate (for example, by collecting ratings for the same items using different participant sample) and the between-items reliability rate (by collecting ratings from the same participants but on different days).

The next chapter examines the results of chapters three through five in light of the goals stipulated in chapter one and the theories discussed in chapter two.

# 6
# Conclusion

## 6.1 Relating the Findings to Theories of Metaphor Processing

The goal of this dissertation was to investigate the way in which the theoretical debate on metaphor processing could be pushed forward by examining previously understudied linguistic phenomena. Crucially, this was done by incorporating the insights and methodology from the field of situated and incremental language processing. This field has shown, for example, that deploying eye-tracking during the inspection of a visual context can resolve broader theoretical issues in cognitive science (Tanenhaus et al., 1995), investigate subtle non-referential links between visual context and sentence processing (Guerra & Knoeferle, 2014, 2017), provide fine-grained evidence regarding the way in which participants derive pragmatic meaning (Huang & Snedeker, 2009), uncover how associated semantic and phonological features play a role during lexical processing (Huettig & Altmann, 2005; Huettig & Hartsuiker, 2008; Tanenhaus et al., 2000), and show that comprehenders use the incrementally-developing sentence interpretation to anticipate upcoming discourse referents (Altmann, 1999; Altmann & Kamide, 2007; Kamide et al., 2003), among other things (see for example Huettig et al., 2011; Knoeferle & Guerra, 2016; Knoeferle, 2016, for systematic reviews).

Specifically, I was interested in the way in which two leading theories of metaphor comprehension - The Implicit Comparison View (Gentner & Bowdle, 2008) and the Category Inclusion View (Glucksberg, 2003; Sperber & Wilson, 2008) - make differentiating predictions on three open issues regarding metaphor comprehension: the issue of symmetry, the role of literal features of the metaphoric vehicle and the effect of potential mediating factors (see Holyoak & Stamenković, 2018, for the most systematic review of these issues to date). The different positions were discussed in detail in chapter two and were broadly summarized in Table 1.1, which is reproduced here as Table 6.1.

**Table 6.1:** Revisiting the theoretical predictions for each empirical issue.

| Empirical Issue | Position of Category Inclusion View | Position of Implicit Comparison View |
|---|---|---|
| *Role of Literal Features* | Suppressed at earliest stages of processing. Should cause interference if pre-activated | Suppressed only in later stages of processing. Should facilitate processing if pre-activated |
| *Symmetry of Processing* | Processing is asymmetric. Reversal of position of the elements should cause difference in earliest stages of processing | Processing is symmetric. Reversal of position of the elements should not cause differences in earliest stages of processing. |
| *Mediating Factors* | Aptness is true mediator of processing mode. | Conventionality or familiarity mediate processing mode. |

The role of the literal features of the metaphoric vehicle was addressed in chapter three. Here, the eye-tracking during reading method was used to investigate how the processing of German verbal metaphors was modulated by preceding exposure to a visual representation of a feature of the literal meaning of the metaphoric verb: Participants saw either video of a ball being trapped by a box (contained condition) or a video of a ball bouncing freely (not-contained condition). They then read metaphoric sentences in which the vehicle was a verb of physical containment, such as *It was difficult fo the journalist to see his opinion be fenced-in after the change in regime.*

As Table 6.1 above illustrates, the role of literal features of a metaphorically used word is disputed: Whereas the Category Inclusion view sees literal features as mostly irrelevant for metaphor processing (McGlone & Manfredi, 2001), the Implicit Comparison View claims that these should be automatically activated during early stages of processing (Gentner et al., 2001). This difference was directly reflected in the predictions of Experiments 1 and 2: If the Category Inclusion view holds, then seeing the “contained” video before reading the metaphorically used verb should hinder comprehension relative to seeing the “not-contained” video first. If, on the other hand, the Implicit Comparison view is correct, we should find that the “contained” video facilitates comprehension relative to the “not-contained” video.

The results of Experiments 1 and 2 where somewhat inconclusive in this regard: They showed neither a facilitation nor an interference effect of the “contained” video on the reading times of the metaphoric verb or of any other sentence region. An initial effect of the matching video found on the response time patterns to questions that followed the metaphoric sentence was shown to be unrelated to metaphor processing in Experiment 3. However, the results of Experiment 4 showed that, in the absence of a sentential context, the same match video used in Experiments 1 and 2 was capable of modulating processing times of the de-contextualized critical verbs: Participants in this experiment showed shorter response times in a lexical decision task when the verbs were preceded by the “contained” compared to when they were preceded by the “not-contained” video.

This pattern of findings suggests, with some caution, that the null-effects found in Experiments 1 and 2 can be interpreted meaningfully: It is possible that the literal features of the metaphorically used verbs that were primed by the “contained” video were simply ignored during processing. This could have been brought about by the fact that these features did not contribute to the construction of metaphoric meaning. When viewed this way, the results could be seen as offering tentative support for the Category Inclusion View: Literal features need not be activated during comprehension of verbal metaphors, even if these features would normally facilitate processing of the same verbs when the verbs are processed without a sentential context.

This could be understood as a consequence of the activated set of parameters that the metaphoric topic is believed to provide during metaphor comprehension: Since the verbal vehicle (*fenced-in*) was always read after the topic (*The journalist's opinion*), it might have been possible for participants to have narrowed down the set of expected upcoming discourse by considering the topic's semantic feature of [-concrete] (i.e. by taking into account the fact that an opinion is an abstract entity). Upon encountering the verb, participants did not integrate the video with the verb because they had previously established that the video was irrelevant for comprehension, given that the semantic features activated by the video (physical containment) were incompatible with the parameters provided by the topic ([- concrete]).

It is possible that the scope of these results is limited to verbal metaphors only, as it has been claimed that verbal and nominal metaphors might undergo distinct processes given their different syntactic attachment structures (Cardillo et al., 2010; Schmidt et al., 2010). However, as discussed in chapter three, there is evidence showing that verbal and nominal metaphors do not differ in the type of neural substrate that they recruit during processing (Cardillo et al., 2012). Additionally, Category Inclusion accounts have been extended in the past to account for verbal metaphors using the same mechanisms believed to be at play during processing of nominal metaphors (Kintsch, 2000; Torreano et al., 2005). Given these empirical and theoretical arguments, which are additionally supported by Ockham's razor, the burden of proof rests on Implicit Comparison theoreticians to show that verbal metaphors are not processed similarly to nominal metaphors.

The second test of the theories, presented in chapter four, dealt with the issue of symmetry: Does altering the relative position of the elements of a metaphor affect the way in which they are processed? The positions of the theories in this regard are also shown in Table 6.1: Implicit Comparison Views see metaphors as being symmetric, meaning that both elements play have the same function, at least during initial stages of processing (Gentner & Wolff, 1997; Wolff & Gentner, 2011).

Category inclusion views, on the other hand, claim that the roles of the elements are different from the earliest stages onward, meaning that changes in the position of the elements should result in changes in the way in which the metaphor is processed (D. Chiappe, Kennedy, & Smykowski, 2003; Glucksberg, 2001, pg. 55-58).

Chapter four addressed this question by examining German verb-object metaphors with a Visual World Paradigm. As opposed to nominal metaphors, verb-object metaphors allow for a natural reversal of the position of the metaphor's elements without rendering the expression infelicitous. The choice of materials was therefore tailored to explore this particular theoretical issue. In the two Experiments of chapter four, participants listened to verb-object metaphors (e.g. *Sebastian feeds a princess (verb-early conditions)/will a princess feed(verb-late conditions) and will relentlessly stand by the cat(metaphoric conditions) /by the noble woman (literal conditions)*) after having read a linguistic context facilitating a literal or a metaphoric interpretation (Experiment 5) or no context whatsoever (Experiment 6). The sentences participants heard in both Experiments were either in the present or future tense (verb-early vs. verb-late conditions), effectively altering the order of presentation of the metaphoric vehicle (*princess*) and the verbal topic (*feeds*). During listening, participants saw 4 images on display: One compatible with a literal interpretation of the metaphoric vehicle (a princess), one compatible with a metaphoric interpretation of the vehicle (a cat), and two distractors.

The results of chapter four support the idea that verb-object metaphors are processed asymmetrically. In Experiment 5, after reading a metaphorically-biasing context, participants directed their gaze preferentially towards the metaphoric picture (the cat) relative to the literal picture (a princess) when hearing the metaphoric vehicle (*princess*) after having heard the verbal topic (*feeds*); but they instead preferred to inspect the literal picture relative to the metaphoric picture when hearing the vehicle before the verbal topic. Crucially, this difference was not found when participants read a literally-biasing context (literal context conditions, Experiment 5), nor when they read no context at all (Experiment 6). This strongly suggests that the difference between viewing behavior can be attributed to differences in the way in which a metaphoric interpretation was constructed.

The pattern of results found in chapter four was predicted and can best be accommodated by Category Inclusion views: when understanding a metaphor, participants will more swiftly grasp the vehicle's metaphoric meaning if they have previously heard the verbal topic compared to when they haven't heard it yet. This can be explicated using Altmann and Kamide (2007)'s linking hypothesis regarding the interpretation of the gaze record in the Visual World Paradigm: When hearing the metaphoric vehicle after the verbal topic (verb-early metaphoric condition), participants determined that the metaphoric picture has the greatest amount of conceptual overlap with the heard word. This suggests that participants might have modulated the meaning of the metaphoric vehicle from *princess* to PRINCESS*, an ad hoc category that includes all humans and animals thought to be spoiled, needy, etc. They were able to do so because they had previously processed the parameters needed for the construction of this ad hoc category when they heard the verbal topic (*feeds* restricts its upcoming object to being an animal or a baby, not an adult human).

When hearing the metaphoric vehicle before the topic (verb-late metaphoric condition), participants were not able to rapidly construct the ad hoc category and instead were likely to derive a literal interpretation of *princess*. Because of this, they determined that the picture of the literal princess had the greatest conceptual overlap with the heard metaphoric vehicle and thus preferentially looked at that picture over the the picture of the cat.

Implicit Comparison views, by contrast, would struggle to accommodate these results: During structural alignment, believed to be a key first stage in processing, it would be necessary to retrieve the full structure of the metaphoric vehicle in order to compare it to the structure of the verbal topic. This should have translated into equal amounts of conceptual overlap between the literal picture (the princess) and the heard vehicle (*princess*) regardless of the vehicle's relative position to the verbal topic. The account would have to be modified to incorporate these findings.

It is hard to make the case that the results from chapter four should not generalize to nominal metaphors. The main claim present in Schmidt et al. (2010) against the comparability of nominal and verbal metaphors, for example, is centered around the processing of the metaphoric vehicle. The authors argue that when a verb is the vehicle, the verb's concrete and sensory features are lost and the verb is abstracted. In a nominal metaphor, on the other hand, they claim that these features are not lost and a process of comparison takes place between topic and vehicle. In verb-object metaphors we do not have a verb as the vehicle, but a noun. Following Schmidt et al. (2010) then, we could in principle assume that the construction of metaphoric meaning occurs similarly for verb-object metaphors as it does for "X is a Y" nominal metaphors. Again, the burden of proof rests on accounts claiming that nominal and verb-object metaphors are processed differently.

Finally, chapter five investigated a third unresolved issue in this debate: The role of potential mediating factors during processing. Specifically, the role of aptness, conventionality and familiarity. Table 6.1 shows the predictions of each set of views in this regard: Indirect comparison views claim that a metaphor can only change from being processed through analogy toward being processed through category inclusion as a function of the vehicle's frequency of association with its metaphoric meaning (Bowdle & Gentner, 2005), which is known as conventionality, or, alternatively, as a function of the frequency of exposure to a metaphoric expression as a whole (Thibodeau & Durgin, 2011), which is known as familiarity. Category inclusion views, on the other hand, believe that any change in processing mode should be a consequence of the way in which the salient features of the vehicle interact with contextually relevant properties of the topic, independently of the frequency of exposure, which is defined as aptness (Glucksberg, 2008; Jones & Estes, 2006).

The investigation in this chapter was exploratory in nature: I collected ratings on all three dimensions (familiarity, aptness and conventionality) for the items used in Experiments 2 and 5. I then re-fitted the statistical models used to analyze the data in those experiments including the newly collected ratings as covariates. The new models, which independently fitted each one of the different ratings, were compared both to a base model and to each other. The results suggest that for both verbal metaphors and verb-object metaphors aptness provided the best fit to the data and was the only factor to show a significant effect on total reading times of the metaphoric vehicle (Experiment 10) and on log-gaze ratios in the time window immediately following the second element of the metaphor (Experiment 14).

Importantly, aptness modulated processing in a way consistent with Category Inclusion views: For the verbal metaphors, more apt metaphors were read faster than less apt ones, localized to the total reading times of the verb itself. For the verb-object metaphors, participants inspected the target metaphoric picture more so for high-apt metaphors than for low-apt metaphors in the time window immediately following the presentation of both metaphoric elements. Both of these findings suggest that aptness - and not conventionality or familiarity - effectively modulated processing so that more apt metaphors were understood faster and more swiftly than less apt metaphors. However, it is important to note that these results are only exploratory and should in future be supported by a confirmatory analysis of newly collected data.

To summarize, the evidence coming from all three empirical chapters appears to be more in line with category inclusion views than with indirect comparison views: Chapter three presents the strongest evidence in this regard, directly showing support for this view's explicit prediction. Chapter two failed to support the strong prediction made by Category inclusion views regarding the role of literal features (see Table 6.1), but it could be seen as supporting a weaker version, in which literal features do not interfere with processing but are instead ignored. Chapter five presents the weakest evidence in this regard, since it consisted of an exploratory analysis, but its results are nevertheless consistent only with Category Inclusion views and not with Implicit Comparison views.

## 6.2 Specifying and Generalizing the Category Inclusion View

Besides providing support for Category Inclusion views, the current dissertation also lays out a road map for further specifications of theoretical accounts of metaphor processing. The main way in which theories of metaphors can be specified based on the outcome of this dissertation concerns the time course of metaphor processing and the point at which effects of aptness on metaphor comprehension should be observable.

In this regard, Experiments 10 and 14 are highly informative: In both cases, an effect of aptness was found in later measures of metaphoric comprehension only: Aptness modulated total reading times of the verb region in Experiment 10 as well as log-gaze ratios in the time region immediately following the second element of the metaphoric expression in Experiment 14. This apparent consistency in timing across paradigms and metaphor types suggests the working hypothesis that a metaphor's aptness comes into play during later processing stages. From a Category Inclusion perspective, this could mean restricting the effect of aptness to the later stages of lexical modulation, that is, only after the relevant parameters for modulation have been set.

The results of Experiment 5 suggest a further way in which Category Inclusion views comes can be refined: When hearing the vehicle in the late-verb metaphoric condition, participants preferred to examine the literal picture suggesting that they might be deriving a literal interpretation of the word *princess* up to this point. However, once they had heard the verbal topic, participants quickly settled on a metaphoric interpretation, as evidenced by the way in which they preferentially inspected the metaphoric picture in every time window that followed the verbal topic (similarly to the way it was done in the early-verb metaphoric condition) as well as by the accuracy in selecting the metaphoric picture at the end of the trial, which did not differ from the early-verb metaphoric condition. This means that category inclusion theories cannot limit themselves to describing the way in which processing occurs when the vehicle follows the topic. They must also account for the way in which processing unfolds when the position of the elements is reversed.

Finally, as shown in chapter three, pre-activating literal features of a metaphor does not have to result in interference with the metaphor comprehension process itself, contrary to what has been predicted by category inclusion models. This means that there is room for theoretical improvement in order to incorporate these results.

A refined version of the Category Inclusion model could work as follows: When language comprehenders first encounter the metaphoric topic, they constrain their expectations regarding upcoming information together with the contextually-derived mental representations that they have developed up to this point.

When they then encounter the metaphoric vehicle, they evaluate the way in which this incoming linguistic information could be integrated with their built-up representations and determine that, given the parameters previously established, they should modulate the meaning of the vehicle in order to construct an ad hoc category. When doing so, language comprehenders can ignore semantic information that is part of the encoded meaning of the metaphoric vehicle but incompatible with the metaphoric interpretation (perhaps up to a particular threshold of salience). Only at this point do they consider the metaphor's aptness, i.e. the degree to which the most salient features of the vehicle can capture relevant properties of the contextually situated metaphoric topic: Metaphors with a higher degree of aptness will go through this stage more quickly than metaphors with a lower degree of aptness.

When language comprehenders first encounter the metaphoric vehicle, they will initially fail to find a way to integrate the word with the previous context. This will result in first deriving a literal interpretation of the vehicle. Once the metaphoric topic is encountered, however, this initial interpretation is revised. This happens because the topic, together with the relevant context, provide narrow-enough parameters at this point in time to revise the initial interpretation of the vehicle and construct the appropriate ad hoc category. This last stage is again sensitive to aptness: The more apt a metaphor is, the faster the final construction of meaning.

This preliminary hypothesis, which could be called the Generalized Category Inclusion View, could be applicable to other types of metaphors, such as nominal or adjectival metaphors. Testing these predictions should be the goal of future investigations in this matter.

## 6.3 The Role of Context During Metaphor Processing

Historically, metaphor research has been focused on testing theories on de-contextualized nominal metaphors. One notable early exception is the work by Gildea and Glucksberg (1983), which followed up on previous research on the automaticity of metaphor processing (Glucksberg et al., 1982). In Glucksberg et al. (1982), the authors had participants give true/false judgments on de-contextualized nominal metaphors (*some jobs are jails*), and literally false (*some jobs are birds*) statements. They found that participants had a harder time stating that metaphors are false compared to the literally false statements, which they dubbed the metaphor interference effect. The authors reasoned that, even though literally false, sentences such as *some jobs are jails* were automatically understood as meaningful metaphors, so that it took participants additional effort to recognize that these were indeed literally false. They noticed, however, that when metaphors seemed to be particularly "bad", no metaphor interference effect was elicited. Would a facilitating context be able to induce such an effect? What would such a context look like? These were the questions asked by Gildea and Glucksberg (1983).

In three experiments, Gildea and Glucksberg (1983) showed participants low-apt metaphors (*a marriage is an icebox*) preceded either by a literal context (*winters are cold*), a metaphoric context (*people are cold*), a literal context that primed the general semantic domain (*summers are warm*) or no context at all. They found that without a context, these low-apt metaphors did not elicit a metaphor interference effect. However, in all three of the context conditions, the metaphors elicited an interference effect similar to the one observed for de-contextualized high-apt metaphors in Glucksberg et al. (1982). The authors conclude that even an incompatible linguistic context (*summers are warm*) can facilitate metaphor comprehension by simply highlighting the relevant parameters needed for the meaning of the metaphoric vehicle to be modulated.

Similarly, the experiments in chapter four of the present dissertation highlight the dramatic way in which context provides the parameters necessary for the construction of metaphoric meaning. In Experiment 6, participants preferentially understood de-contextualized verb-object utterances in a literal way: When hearing the object of the verb (*Prinzessin*), they preferentially inspected the picture of the literal princess over the metaphoric picture regardless of the object's position in the sentence, and they more often selected the literal image as the appropriate referent at the end of the trial. By contrast, in the presence of a linguistic context (Experiment 5), the object can be easily understood as a metaphoric vehicle, to the point where participants preferentially inspected the metaphoric picture over the literal picture for the duration of the entire time window (Experiment 5, vehicle region, verb-early metaphoric condition).

This stark difference in gaze patterns between Experiments 5 and 6 can be understood as following the same general principles put forth by Gildea and Glucksberg (1983): When a metaphorical vehicle is not given enough constraining parameters for its meaning to be modulated, language users will probably fail to do so and instead fall back on a conventional literal interpretation. When provided with an appropriate context, the set of parameters needed for modulation is sharpened by making the dimension more salient, effectively turning a low-apt metaphor into a high-apt metaphor.

It's important to note that there is no need to postulate a metaphor-specific mechanism regarding the way in which context facilitates comprehension: Instead, The way in which a linguistic context facilitates metaphoric meaning comprehension can be seen to be similar to the way in which context modulates language comprehension in other types of tasks. Linguistic contexts have been shown to rapidly override default syntactic preferences (Altmann et al., 1998), generate expectations about specific upcoming nouns (e.g. Van Berkum et al., 2005), modulate sense selection during word recognition (Sereno et al., 2003), as well as ease processing and recognition of other instances of figurative language such as irony (Rodriguez Ronderos, Noveck, et al., 2020), among many other things.

The results of Experiments 1 and 2, on the other hand, show the limitations of the way in which context (in this case visual) interacts with metaphor processing. Though both rivaling theories would predict that priming literal features should have an effect on metaphor processing (facilitation effect, in the case of Implicit Comparison views, and interference effect, in the case of category inclusion views), the results of Experiments 1 and 2 showed a null-effect of the visual context on processing the metaphoric sentence. This is somewhat puzzling, considering how the same videos facilitated responses to the verbs present in the sentences in a lexical decision task (Experiment 4) and how effects of the visual context on subsequent sentence reading have been found on multiple occasions in the past (e.g. Guerra & Knoeferle, 2014; Guerra & Knoeferle, 2017; Guerra & Knoeferle, 2018). Experiment 2, in particular, was sufficiently powered to detect any such subtle effects of the visual context.

Two possible explanations for these null-effects come to mind. The first one is the one given earlier in this chapter and in the interim conclusion of chapter three: Participants could have ignored the information derived from the visual prime in Experiments 1 and 2 because it was unnecessary for the construction of metaphoric meaning. In Experiment 4, this information elicited a priming effect because no metaphoric meaning was constructed.

The second possible explanation relates to the differences in the set-up of Experiments 1, 2 and 4. In Experiment 4 (where effects of the visual context were found on the processing of the isolated verb), participants saw only two critical items: One in the match and one in the mismatch condition. In Experiments 1 and 2, participants saw the full set of 36 items, with half of them preceded by the matching and the other half by the mismatching video prime. It is possible then, than participants in Experiments 1 and 2 became desensitized to the information provided by the visual primes because of repeated exposure (the visual primes were the same for all critical items), thus masking a possible, underlying true effect (be it of interference or facilitation).

This second interpretation, though compatible with the pattern of results, seems less likely if we consider that the Experiments in Guerra and Knoeferle (2014) and Guerra and Knoeferle (2018), for example, also used the same visual primes for all items and were nevertheless able to detect both facilitation and interference effects. However, deciding between these two interpretations of the results is ultimately an empirical question that requires further follow-up studies.

## 6.4 Beyond Metaphors: Interaction of the Linguistic Context and a Verb's Semantic Features

Another important finding of this dissertation pertains to the way in which information derived from a linguistic context interacts with the information assumed to be part of the semantics of a verb, independently of the presence of a metaphoric expression.

Previous studies have shown that in the absence of a context, participants use the information obtained from a verb's selectional restrictions to anticipate an upcoming referent (e.g. Altmann & Kamide, 1999; Hintz et al., 2017). Furthermore, when a verb (*ride*) on its own does not provide enough restrictions to select between two possible visual referents (a motorcycle and a carousel), participants incorporate information about the verb's stated agent (*a man* or *a girl*) in order to anticipate the likely object upon hearing the verb (Kamide et al., 2003). But what happens when the contextual information and the verb's selectional restrictions point towards different possible upcoming verbal objects?

This was investigated in Experiment 5, where participants read a context biasing towards a literal interpretation of an upcoming spoken utterance (in the literal context bias conditions). The spoken utterance contained a verb (*füttern*, "feeds") whose selectional restrictions (*füttern* can only take an animal or a baby as its object) were more compatible with the metaphoric picture (a cat) than with the literal picture (a princess), even though the literal picture was the most compatible with the contextually-derived representation (The linguistic context is about Sebastian having a noble woman as a friend). When participants heard the verb, they did not anticipate any upcoming object in particular, judging by the way in which the log-gaze ratio of looks to the metaphoric divided by looks to the literal pictured hovered around zero for the entirety of the region. Participants were uncertain about which object to anticipate: The one most related to the context (the princess) or the one most related to the verb (the cat).

In the metaphoric early-verb conditions of Experiment 5, participants did not display this hesitation when hearing the verb: In this case, they were able to anticipate the object that was both compatible with context and compatible with the verb's selectional restrictions (i.e. the cat). This unequivocal anticipation of the upcoming object is what eventually lead participants to process the metaphoric vehicle with more ease: Hearing the verb before the vehicle in the metaphoric conditions resulted in participants preferentially directing their gaze to the metaphoric picture when hearing the metaphoric vehicle.

It is interesting to compare these results with those of Experiment 6: Here, in the absence of a linguistic context, participants could only rely on the verb's selectional restrictions to anticipate the upcoming object. However, this anticipation effect was only visible in the second half of the verb region, as well as in the comparison between looks to the metaphoric picture and looks to either one of the distractor pictures.

As a whole, these results can be interpreted as meaning that a verb's selectional preferences on their own can facilitate anticipation, but only after some time has passed, i.e. when processing of the verb is already on its way. When the verb's selectional restrictions are coupled with a supporting linguistic context (as was the case of Experiment 5, in the early-verb metaphoric condition), anticipation of the object happened sooner and had an overall larger effect.

This suggests that post-verbal object anticipation effects in the Visual World Paradigm are the result of a complex parallel integration of all of the information leading up to the verb: The richer the mental representation up to the onset of the verb, the easier it will be to anticipate a referent. When only a verb's selectional restrictions are available, anticipation occurs but in a somewhat delayed fashion (Verb region, Experiment 6) compared to when a linguistic context supports the verb's selectional restrictions (Verb region, early-verb metaphoric condition, Experiment 5). When different sources of information collide (such as context and a verb's selectional preferences) anticipation might not be possible at all (Verb region, early-verb literal condition, Experiment 5).

This interpretation is broadly compatible with the view stated by Metusalem et al. (2012), who tested the influence of salient event representations on the anticipation of semantically incompatible words. In an EEG study, they had participants read a discourse (e.g. about a blizzard and kids playing in the snow) followed by a sentence (*the kids built a...*) that ended in a word compatible with the verb's selectional restrictions (*snowman*) or an incompatible one (*towel* or *jacket*). They measured event-related brain potentials time-locked to the onset of the last word. They observed that, when the semantically inappropriate word was overall compatible with the event structure provided in the previous written discourse (*jacket*), it elicited an N400 effect larger than the one elicited by the expected word (*snowman*) but smaller compared to a word that was incompatible with both the written discourse and the verb's selectional restrictions (*towel*). The authors interpreted this finding as suggesting that the written discourse triggered the construction of an event representation that was capable of accommodating the processing of a word despite the word being semantically inappropriate. This interpretation can be extended to the findings of chapter four of this dissertation: The event representation built up to the point of hearing the verb *füttert* in the early-verb literal condition leads to increased activation of the event-related object (the princess) in parallel to the activation of the object compatible with the verb's lexical semantics (the cat). This simultaneous activation leads participants to distribute their visual attention to both objects similarly throughout the region and prevents them from anticipating one object over the other. This results in them having to wait until more linguistic information is available before they can settle on a single object as the preferred referent.

This "wait-and-see" strategy (displayed by participants during the verb region in the early-verb literal condition of Experiment 5) can be accounted for by appealing to the architecture of predictive pre-activation during language processing posited by Kuperberg and Jaeger (2016). Kuperberg and Jaeger (2016) suggest that language comprehenders trade off the benefits and costs of predicting upcoming language by rationally allocating their available metabolic and cognitive resources. The authors formalize rationality in terms of a utility function that weighs the advantages and disadvantages of anticipating information during comprehension. One way in which comprehenders do this, according to Kuperberg and Jaeger (2016), is by estimating the reliability of prior knowledge and using this estimate to calibrate the degree to which they should update their beliefs. In other words, comprehenders assess the level of prediction error before anticipating upcoming input. In the specific case of Experiment 5 of this dissertation, this could mean that participants determined that the input (during the verb region in the early-verb literal condition) had low reliability for anticipation purposes: The verb's semantic information was at odds with the previously read linguistic context. Participants thus maximized their utility by not anticipating any representation and instead waiting until more information became available in order to update their beliefs.

## 6.5 Quo Vadis, Metaphor Research?

With the end of this dissertation in sight, it's important to reflect on the impact that this work might have on future research on metaphor processing.

First and foremost, the present work has shown the need to go beyond the horizon of English nominal metaphors so that the field can begin to solve longstanding theoretical debates. Only by doing this can we start to get a grasp on the phenomenon as a whole. This work is already underway: Bambini et al. (2019), for example, provided preliminary evidence on the processing of Italian literary metaphors, and Coulson et al. (2015) investigated the processing of English adjectival metaphors. In the future, the field would benefit from intensifying efforts in this direction.

Secondly, the Experiments presented in this work appear to support Category Inclusion views more than Implicit Comparison views. Given the interest of the field in resolving this debate, it would be necessary to determine how generalizable these results are by attempting to replicate them and to determine the precise conditions under which they are valid. This is particularly the case for the results of chapter three, in which the interpretation is based on a null result, and chapter five, in which the main analyses were exploratory in nature.

Besides trying to replicate the results, future research should also focus on developing computational psychological models of metaphor comprehension. Formalizing verbal theories, as has been argued by Gervais (2020), Guest and Martin (2020), and van Rooij and Baggio (2020), among others, allows researchers to be certain of the full set of assumptions and commitments that any stated theory has. This can in turn lead to future ways of testing between theories and refining our beliefs concerning the cognitive mechanisms involved in metaphor comprehension.

Finally, different kinds of tests of the theories are necessary in order to definitively be able to choose one theory over the other. One possible avenue of exploration, for example, could be to take a contrastive look at the processing of metaphors and other types of phenomena believed to occur via indirect comparison (such as analogies), and phenomena believed to occur via the construction of ad hoc categories (such as hyperboles). This could lead to uncovering similarities and differences between these processes that would allow for a refinement of a Generalized Category Inclusion View. Following this path will bring us to a deeper understanding of the way in which metaphors are understood and how the construction of metaphoric meaning relates to other cognitive processes.

# Appendices

# A

# Appendix A

## A.1 Critical Items for Experiments 1-3

**Table A.1:** critical items of Experiments 1-3

| **“item number”** | **“item, Experiment 1”** | **“item, Experiment 2”** | **“item, Experiment 3”** |
|---|---|---|---|
| 1 | “Es war für den Polizisten schwierig, seinen Aufstiegsdrang nach dem Korruptionsvorwurf gedeckelt zu sehen.” | “Dass sein Aufstiegsdrang gedeckelt wurde nach dem Korruptionsvorwurf, war schwierig für den Polizisten.” | “Dass sein Aufstiegsdrang beneidet wurde nach dem Korruptionsvorwurf, war schwierig für den Polizisten.” |
| 2 | “Es war für den Auszubildenden schwierig, seine Ambitionen nach dem Rechenfehler gedeckelt zu sehen.” | " Dass seine Ambitionen gedeckelt wurden nach dem Rechenfehler, war schwierig für den Auszubildenden." | “Dass seine Ambitionen beneidet wurden nach dem Rechenfehler, war schwierig für den Auszubildenden.” |
| 3 | “Es war für den Premierminister schwierig, seinen Einfluss nach der Wahlschlappe ummauert zu sehen.” | " Dass sein Einfluss ummauert wurde nach der Wahlschlappe, war schwierig für den Premierminister." | “Dass sein Einfluss unterschätz wurde nach der Wahlschlappe, war schwierig für den Premierminister.” |

| **“item number”** | **“item, Experiment 1”** | **“item, Experiment 2”** | **“item, Experiment 3”** |
|---|---|---|---|
| 4 | “Es war für den Gruppenleiter schwierig, seine Unabhängigkeit nach dem Betrugsfall ummauert zu sehen.” | “Dass seine Unabhängigkeit ummauert wurde nach dem Betrugsfall, war schwierig für den Gruppenleiter.” | “Dass seine Unabhängigkeit erörtert wurde nach dem Betrugsfall, war schwierig für den Gruppenleiter.” |
| 5 | “Es war für den Moderator schwierig, seine Experimentierfreude nach dem Quotensturz gefesselt zu sehen.” | “Dass seine Experimentierfreude gefesselt wurde nach dem Quotensturz, war schwierig für den Moderator.” | “Dass seine Experimentierfreude vernachlässigt wurde nach dem Quotensturz, war schwierig für den Moderator.” |
| 6 | “Es war für den Arzt schwierig, seine Ermessensfreiheit nach der Fehleinschätzung umrandet zu sehen.” | “Dass seine Ermessensfreiheit umrandet wurde nach der Fehleinschätzung, war schwierig für den Arzt.” | “Dass seine Ermessensfreiheit erörtert wurde nach der Fehleinschätzung, war schwierig für den Arzt.” |
| 7 | “Es war für den Rentner schwierig, seine Selbstständigkeit nach der Diagnose umrandet zu sehen.” | “Dass seine Selbstständigkeit umrandet wurde nach der Diagnose, war schwierig für den Rentner.” | “Dass seine Selbstständigkeit angezweifelt wurde nach der Diagnose, war schwierig für den Rentner.” |
| 8 | “Es war für den Schüler schwierig, seine Freizeit nach der Zeugnisvergabe umwickelt zu sehen.” | “Dass seine Freizeit umwickelt wurde nach der Zeugnisvergabe, war schwierig für den Schüler.” | “Dass seine Freizeit vergessen wurde nach der Zeugnisvergabe, war schwierig für den Schüler.” |
| 9 | “Es war für den Stürmer schwierig, seine Trainingszeit nach der Verletzung umgrenzt zu sehen.” | “Dass seine Trainingszeit umgrenzt wurde nach der Verletzung, war schwierig für den Stürmer.” | “Dass seine Trainingszeit besprochen wurde nach der Verletzung was schwierig für den Stürmer.” |
| 10 | “Es war für den Studenten schwierig, sein Mitbestimmungsrecht nach den Protesten umgrenzt zu sehen.” | “Dass sein Mitbestimmungsrecht umgrenzt wurde nach den Protesten, war schwierig für den Studenten.” | “Dass sein Mitbestimmungsrecht aberkannt wurde nach den Protesten, war schwierig für den Studenten.” |

| **“item number”** | **“item, Experiment 1”** | **“item, Experiment 2”** | **“item, Experiment 3”** |
|---|---|---|---|
| 11 | “Es war für den Soldaten schwierig, seinen Optimismus nach der Niederlage umgittert zu sehen.” | “Dass sein Optimismus umgittert wurde nach der Niederlage, war schwierig für den Soldaten.” | “Dass sein Optimismus bemitleidet wurde nach der Niederlage, war schwierig für den Soldaten.” |
| 12 | “Es war für den Redakteur schwierig, seine Meinung nach dem Regimewechsel umgittert zu sehen.” | “Dass seine Meinung umgittert wurde nach dem Regimewechsel, war schwierig für den Redakteur.” | “Dass seine Meinung angezweifelt wurde nach dem Regimewechsel, war schwierig für den Redakteur.” |
| 13 | “Es war für den Staranwalt schwierig, seine Siegesgewissheit nach der Beweisaufnahme umzäunt zu sehen.” | “Dass seine Siegesgewissheit umzäunt wurde nach der Beweisaufnahme, war schwierig für den Staranwalt.” | “Dass seine Siegesgewissheit besprochen wurde nach der Beweisaufnahme, war schwierig für den Staranwalt.” |
| 14 | “Es war für den Lehrer schwierig, seine Privilegien nach der Bildungsreform umzäunt zu sehen.” | “Dass seine Privilegien umzäunt wurden nach der Bildungsreform, war schwierig für den Lehrer.” | “Dass seine Privilegien beklagt wurden nach der Bildungsreform, war schwierig für den Lehrer.” |
| 15 | “Es war für den Bildhauer schwierig, seine Entwicklung nach der Akademieschließung beengt zu sehen.” | “Dass seine Entwicklung beengt wurde nach der Akademieschließung, war schwierig für den Bildhauer.” | “Dass seine Entwicklung bewundert wurde nach der Akademieschließung, war schwierig für den Bildhauer.” |
| 16 | “Es war für den Helden schwierig, seine Kraft nach dem Verkehrsunfall beengt zu sehen.” | “Dass seine Kraft beengt wurde nach dem Verkehrsunfall, war schwierig für den Helden.” | “Dass seine Kraft bewundert wurde nach dem Verkehrsunfall, war schwierig für den Helden.” |
| 17 | “Es war für den Investor schwierig, seine Expansionspläne nach dem Börsenkrach eingeschnürt zu sehen.” | “Dass seine Expansionspläne eingeschnürt wurden nach dem Börsenkrach, war schwierig für den Investor.” | “Dass seine Expansionspläne aufgeschoben wurden nach dem Börsenkrach, war schwierig für den Investor.” |

| **“item number”** | **“item, Experiment 1”** | **“item, Experiment 2”** | **“item, Experiment 3”** |
|---|---|---|---|
| 18 | “Es war für den Musiker schwierig, seinen Freigeist nach der Konzertpleite eingeschnürt zu sehen.” | “Dass sein Freigeist eingeschnürt wurde nach der Konzertpleite, war schwierig für den Musiker.” | “Dass sein Freigeist beleidigt wurde nach der Konzertpleite, war schwierig für den Musiker.” |
| 19 | “Es war für den Grundbesitzer schwierig, seine Kauflust nach der Revolution eingekapselt zu sehen.” | “Dass seine Kauflust eingekapselt wurde nach der Revolution, war schwierig für den Grundbesitzer.” | “Dass seine Kauflust verachtet wurde nach der Revolution, war schwierig für den Grundbesitzer.” |
| 20 | “Es war für den Künstler schwierig, seine Gestaltungsfreiheit nach der Zeitungskritik eingekapselt zu sehen.” | “Dass seine Gestaltungsfreiheit eingekapselt wurde nach der Zeitungskritik, war schwierig für den Künstler.” | “Dass seine Gestaltungsfreiheit beleidigt wurde nach der Zeitungskritik, war für den Künstler schwierig.” |
| 21 | “Es war für den Priester schwierig, seine Autonomie nach den Skandalen einbetoniert zu sehen.” | “Dass seine Autonomie einbetoniert wurde nach den Skandalen, war schwierig für den Priester.” | “Dass seine Autonomie diskutiert wurde nach den Skandalen, war schwierig für den Priester.” |
| 22 | “Es war für den General schwierig, seine Befugnisse nach dem Militärputsch einbetoniert zu sehen.” | “Dass seine Befugnisse einbetoniert wurden nach dem Militärputsch, war schwierig für den General.” | “Dass seine Befugnisse ignoriert wurden nach dem Militärputsch, war schwierig für den General.” |
| 23 | “Es war für den Abgeordneten schwierig, seinen Gesetzentwurf nach der Staatskrise eingeklemmt zu sehen.” | “Dass sein Gesetzentwurf eingeklemmt wurde nach der Staatskrise, war schwierig für den Abgeordneten.” | “Dass sein Gesetzentwurf aufgeschoben wurde nach der Staatskrise, war für den Abgeordneten schwierig.” |

| "item number" | "item, Experiment 1" | "item, Experiment 2" | "item, Experiment 3" |
|---|---|---|---|
| 24 | "Es war für den Professor schwierig, sein Forschungsvorhaben nach der Umstrukturierung eingeklemmt zu sehen." | "Dass sein Forschungsvorhaben eingeklemmt wurde nach der Umstrukturierung, war schwierig für den Professor." | "Dass sein Forschungsvorhaben thematisiert wurde nach der Umstrukturierung, war schwierig für den Professor." |
| 25 | "Es war für den Opernsänger schwierig, sein Selbstvertrauen nach dem Saison-Fehlstart festgekettet zu sehen." | "Dass sein Selbstvertrauen festgekettet wurde nach dem Saison-Fehlstart, war schwierig für den Opernsänger." | "Dass sein Selbstvertrauen überrumpelt wurde nach dem Saison-Fehlstart, war schwierig für den Opernsänger." |
| 26 | "Es war für den Architekten schwierig, seine Schöpferkraft nach der Budgetkürzung festgekettet zu sehen." | "Dass seine Schöpferkraft festgekettet wurde nach der Budgetkürzung, war schwierig für den Architekten." | "Dass seine Schöpferkraft ignoriert wurde nach der Budgetkürzung, war schwierig für den Architekten." |
| 27 | "Es war für den Anleger schwierig, seine Risikofreude nach dem Kurssturz festgebunden zu sehen." | "Dass seine Risikofreude festgebunden wurde nach dem Kurssturz, war schwierig für den Anleger." | "Dass seine Risikofreude thematisiert wurde nach dem Kurssturz, war schwierig für den Anleger." |
| 28 | "Es war für den König schwierig, seine Macht nach der Verfassungsänderung festgebunden zu sehen." | "Dass seine Macht festgebunden wurde nach der Verfassungsänderung, war schwierig für den König." | "Dass seine Macht belächelt wurde nach der Verfassungsänderung, war schwierig für den König." |
| 29 | "Es war für den Vermieter schwierig, seine Ansprüche nach der Mietrechtsreform abgesteckt zu sehen." | "Dass seine Ansprüche abgesteckt wurden nach der Mietrechtsreform, war schwierig für den Vermieter." | "Dass seine Ansprüche diskutiert wurden nach der Mietrechtsreform, war schwierig für den Vermieter." |

| “item number” | “item, Experiment 1” | “item, Experiment 2” | “item, Experiment 3” |
|---|---|---|---|
| 30 | “Es war für den Unternehmensleiter schwierig, sein Vermächtnis nach der Verstaatlichung abgesteckt zu sehen.” | “Dass sein Vermächtnis abgesteckt wurde nach der Verstaatlichung, war schwierig für den Unternehmensleiter.” | “Dass sein Vermächtnis besteuert wurde nach der Verstaatlichung, war schwierig für den Unternehmensleiter.” |
| 31 | “Es war für den Clown schwierig, seinen Humor nach der Panne eingedämmt zu sehen.” | “Dass sein Humor eingedämmt wurde nach der Panne, war schwierig für den Clown.” | “Dass sein Humor bemitleidet wurde nach der Panne, war schwierig für den Clown.” |
| 32 | “Es war für den Ganoven schwierig, seine Überlegenheit nach der Razzia eingedämmt zu sehen.” | “Dass seine Überlegenheit eingedämmt wurde nach der Razzia, war schwierig für den Ganoven.” | “Dass seine Überlegenheit unterschätzt wurde nach der Razzia, war schwierig für den Ganoven.” |
| 33 | “Es war für den Reporter schwierig, seine Recherchemöglichkeiten nach dem Berufsverbot abgeriegelt zu sehen.” | “Dass seine Recherchemöglichkeiten abgeriegelt wurden nach dem Berufsverbot, war schwierig für den Reporter.” | “Dass seine Recherchemöglichkeiten misstraut wurden nach dem Berufsverbot, war schwierig für den Reporter.” |
| 34 | “Es war für den Bauern schwierig, seine Förderungschancen nach der Gesetzesänderung abgeriegelt zu sehen.” | “Dass seine Förderungschancen abgeriegelt wurden nach der Gesetzesänderung, war schwierig für den Bauern.” | “Dass seine Förderungschancen benachteiligt wurden nach der Gesetzesänderung, war schwierig für den Bauern.” |
| 35 | “Es war für den Bürgermeister schwierig, seine Karriere nach dem Misstrauensantrag abgesperrt zu sehen.” | “Dass seine Karriere abgesperrt wurde nach dem Misstrauensantrag, war schwierig für den Bürgermeister.” | “Dass seine Karriere beendet wurde nach dem Misstrauensantrag, war schwierig für den Bürgermeister.” |

| **"item number"** | **"item, Experiment 1"** | **"item, Experiment 2"** | **"item, Experiment 3"** |
|---|---|---|---|
| 36 | "Es war für den Oppositionellen schwierig, seine Zuversicht nach der Parteispaltung abgesperrt zu sehen." | "Dass seine Zuversicht abgesperrt wurde nach der Parteispaltung, war schwierig für den Oppositionellen." | "Dass seine Zuversicht belächelt wurde nach der Parteispaltung, war schwierig für den Oppositionellen." |

## A.2 Critical Items for Experiment 4

**Table A.2:** critical items of Experiment 4

| **"item number"** | **"target word, Experiment 4"** |
|---|---|
| 1 | "gedeckelt" |
| 2 | "beengt" |
| 3 | "ummauert" |
| 4 | "gefesselt" |
| 5 | "umrandet" |
| 6 | "umwickelt" |
| 7 | "umgrenzt" |
| 8 | "umgittert" |
| 9 | "umzäunt" |
| 10 | "eingeschnürt" |
| 11 | "eingekapselt" |
| 12 | "einbetoniert" |
| 13 | "eingeklemmt" |
| 14 | "festgekettet" |
| 15 | "festgebunden" |
| 16 | "abgesteckt" |
| 17 | "abgeriegelt" |
| 18 | "abgesperrt" |

# B

# Appendix B

## B.1 Critical Contexts for Experiment 5

**Table B.1:** critical linguistic contexts of Experiment 5

| **“item number”** | **“Context”** | **“context bias”** |
|---|---|---|
| 1 | “Sarah und ihre Arbeitskollegen betreten das Büro. Ein unkontrollierter Angestellter ist sehr laut, und droht ohne Grund einen Azubi zu feuern. Sarah wird sich dem Angestellten in den Weg stellen, weil er den Azubi sonst zugrunde richten wird. Für ihr Handeln wird sie danach bejubelt werden.” | “met.” |
| 1 | “Sarah und ihre Arbeitskollegen betreten die Straße. Ein unkontrolliertes Raupenfahrzeug ist laut und droht ohne Grund eine Schule zu rammen. Sarah wird sich dem Raupenfahrzeug in den Weg stellen, weil es die Schule sonst zugrunde richten wird. Für ihr Handeln wird sie danach bejubelt werden.” | “literal” |
| 1 | “Sarah und ihre Arbeitskollegen betreten das Büro. Ein unkontrollierter Angestellter ist sehr laut, und droht ohne Grund einen Azubi zu feuern. Sarah wird sich dem Angestellten in den Weg stellen, weil er den Azubi sonst zugrunde richten wird. Für ihr Handeln wird sie danach bejubelt werden.” | “met.” |
| 1 | “Sarah und ihre Arbeitskollegen betreten die Straße. Ein unkontrolliertes Raupenfahrzeug ist laut und droht ohne Grund eine Schule zu rammen. Sarah wird sich dem Raupenfahrzeug in den Weg stellen, weil es die Schule sonst zugrunde richten wird. Für ihr Handeln wird sie danach bejubelt werden.” | “literal” |

| "item number" | "Context" | "context bias" |
|---|---|---|
| 2 | "Michael geht mit seinen Freunden in die Berge. Dort sieht er eine schlafende, mythische Feuerechse, die den Ruf hat, böse zu sein und Kinder zu fressen. Er versucht leise zu sein, aber sein Handy klingelt und seine Freunde schreien. Das wird sich als Fehler erweisen." | "literal" |
| 2 | "Michael geht mit seinen Freunden in die Bibliothek. Dort sieht er eine schlafende Bibliothekarin, die den Ruf hat, sehr böse zu sein und Kinder zu hassen. Er versucht leise zu sein, aber sein Handy klingelt und seine Freunde lachen. Das wird sich als Fehler erweisen." | "met." |
| 2 | "Michael geht mit seinen Freunden in die Berge. Dort sieht er eine schlafende, mythische Feuerechse, die den Ruf hat, böse zu sein und Kinder zu fressen. Er versucht leise zu sein, aber sein Handy klingelt und seine Freunde schreien. Das wird sich als Fehler erweisen." | "literal" |
| 2 | "Michael geht mit seinen Freunden in die Bibliothek. Dort sieht er eine schlafende Bibliothekarin, die den Ruf hat, sehr böse zu sein und Kinder zu hassen. Er versucht leise zu sein, aber sein Handy klingelt und seine Freunde lachen. Das wird sich als Fehler erweisen." | "met." |
| 3 | "Thomas besucht seine Familie und begegnet dabei seiner gefühllosen, kalten Oma. Sie hat Schwierigkeiten, Liebe zu verstehen. Er wird mit ihr reden und Gefühle zeigen. Er hofft, sie zu erwärmen, aber es wird nicht funktionieren." | "met." |
| 3 | "Thomas besucht ein technisches Museum und begegnet einer intelligenten Maschine. Sie hat Schwierigkeiten, Liebe zu verstehen. Er wird mit ihr reden und Gefühle zeigen. Er hofft, sie zu erwärmen, aber es wird nicht funktionieren." | "literal" |
| 3 | "Thomas besucht seine Familie und begegnet dabei seiner gefühllosen, kalten Oma. Sie hat Schwierigkeiten, Liebe zu verstehen. Er wird mit ihr reden und Gefühle zeigen. Er hofft, sie zu erwärmen, aber es wird nicht funktionieren." | "met." |
| 3 | "Thomas besucht ein technisches Museum und begegnet einer intelligenten Maschine. Sie hat Schwierigkeiten, Liebe zu verstehen. Er wird mit ihr reden und Gefühle zeigen. Er hofft, sie zu erwärmen, aber es wird nicht funktionieren." | "literal" |
| 4 | "Karin wird demnächst eine Pflegeeinrichtung leiten. Obwohl sie sehr optimistisch ist, weiß sie, dass die Senioren deprimiert sind und die Bezahlung sehr schlecht ist. Sie sind auch noch ziemlich trostlos und alt. Sie freut sich aber trotzdem auf die Arbeit." | "literal" |
| 4 | "Karin wird demnächst eine Fußballmannschaft leiten. Obwohl sie sehr optimistisch ist, weiß sie, dass die Fußballspieler unmotiviert sind und deren Leistung schlecht ist. Sie sind auch noch sehr träge und langsam. Sie freut sich aber trotzdem auf die Arbeit." | "met." |

| “item number” | “Context” | “context bias” |
|---|---|---|
| 4 | “Karin wird demnächst eine Pflegeeinrichtung leiten. Obwohl sie sehr optimistisch ist, weiß sie, dass die Senioren deprimiert sind und die Bezahlung sehr schlecht ist. Sie sind auch noch ziemlich trostlos und alt. Sie freut sich aber trotzdem auf die Arbeit.” | “literal” |
| 4 | “Karin wird demnächst eine Fußballmannschaft leiten. Obwohl sie sehr optimistisch ist, weiß sie, dass die Fußballspieler unmotiviert sind und deren Leistung schlecht ist. Sie sind auch noch sehr träge und langsam. Sie freut sich aber trotzdem auf die Arbeit.” | “met.” |
| 5 | “Andreas arbeitet in der Steuerbehörde und besucht beruflich ein Flüchtlingslager. Er kommt pünktlich an, und spricht über Finanzen mit einer Nonne, welche die Rechte der Flüchtlinge Namibias verteidigt. Er will der frommen Frau helfen, weil sie den ganzen Tag arbeitet und die Freude dabei nicht verliert. Er wird sich intensiv damit auseinandersetzen.” | “met.” |
| 5 | “Andreas arbeitet in der Steuerbehörde und besucht privat ein Kunstmuseum. Er ist betrunken, und spricht aus Versehen über Finanzen mit einem Gemälde, welches die Verkündung der Geburt Jesu darstellt. Er will dem Götterboten helfen, weil er die frohe Botschaft verkündet und die Ruhe dabei nicht verliert. Er wird sich intensiv damit auseinandersetzen.” | “literal” |
| 5 | “Andreas arbeitet in der Steuerbehörde und besucht beruflich ein Flüchtlingslager. Er kommt pünktlich an, und spricht über Finanzen mit einer Nonne, welche die Rechte der Flüchtlinge Namibias verteidigt. Er will der frommen Frau helfen, weil sie den ganzen Tag arbeitet und die Freude dabei nicht verliert. Er wird sich intensiv damit auseinandersetzen.” | “met.” |
| 5 | “Andreas arbeitet in der Steuerbehörde und besucht privat ein Kunstmuseum. Er ist betrunken, und spricht aus Versehen über Finanzen mit einem Gemälde, welches die Verkündung der Geburt Jesu darstellt. Er will dem Götterboten helfen, weil er die frohe Botschaft verkündet und die Ruhe dabei nicht verliert. Er wird sich intensiv damit auseinandersetzen.” | “literal” |
| 6 | “Stefan ist Journalist und fährt nach Afrika, um das Thema Tierquälerei zu recherchieren. Als er da ist, findet er in der Wüste ein ungewöhnliches Raubtier, das stark misshandelt und sogar angeschossen wurde. Stefan kommt ihm näher und wird das traurige Stöhnen des Tieres auf Tonband aufnehmen. Diese Begegnung wird ihm lange im Gedächtnis bleiben.” | “literal” |

| **"item number"** | **"Context"** | **"context bias"** |
|---|---|---|
| 6 | "Stefan ist Journalist und fährt nach Frankfurt, um das Thema Steuerbetrug zu recherchieren. Als er da ist, findet er in der Bank einen amoralischen Makler, der den Staat betrügt und sogar Geldwäscherei betreibt. Stefan kommt ihm näher und wird das Angeben des Kapitalisten auf Tonband aufnehmen. Diese Begegnung wird ihm lange im Gedächtnis bleiben." | "met." |
| 6 | "Stefan ist Journalist und fährt nach Afrika, um das Thema Tierquälerei zu recherchieren. Als er da ist, findet er in der Wüste ein ungewöhnliches Raubtier, das stark misshandelt und sogar angeschossen wurde. Stefan kommt ihm näher und wird das traurige Stöhnen des Tieres auf Tonband aufnehmen. Diese Begegnung wird ihm lange im Gedächtnis bleiben." | "literal" |
| 6 | "Stefan ist Journalist und fährt nach Frankfurt, um das Thema Steuerbetrug zu recherchieren. Als er da ist, findet er in der Bank einen amoralischen Makler, der den Staat betrügt und sogar Geldwäscherei betreibt. Stefan kommt ihm näher und wird das Angeben des Kapitalisten auf Tonband aufnehmen. Diese Begegnung wird ihm lange im Gedächtnis bleiben." | "met." |
| 7 | "Susanne geht zum ersten Mal in den Reichstag. Es gibt einen Empfang mit dem einzigen noch lebenden Politiker aus der vorletzten Republik. Sie wird in den großen Raum hineingehen und dem sehr alten Mann die Hand reichen. Danach wird sie lange von diesem Moment schwärmen." | "met." |
| 7 | "Susanne geht zum ersten Mal ins Naturkunde Museum. Es gibt einen Raum mit dem einzigen noch erhaltenen Riesenskelett aus der kleinen Eiszeit. Sie wird in den großen Raum hineingehen und dem alten Skelett Respekt zollen. Danach wird sie lange von diesem Moment schwärmen." | "literal" |
| 7 | "Susanne geht zum ersten Mal in den Reichstag. Es gibt einen Empfang mit dem einzigen noch lebenden Politiker aus der vorletzten Republik. Sie wird in den großen Raum hineingehen und dem sehr alten Mann die Hand reichen. Danach wird sie lange von diesem Moment schwärmen." | "met." |
| 7 | "Susanne geht zum ersten Mal ins Naturkunde Museum. Es gibt einen Raum mit dem einzigen noch erhaltenen Riesenskelett aus der kleinen Eiszeit. Sie wird in den großen Raum hineingehen und dem alten Skelett Respekt zollen. Danach wird sie lange von diesem Moment schwärmen." | "literal" |
| 8 | "Martin und seine Freundin haben zusammen eine Villa gekauft. Sie waren überglücklich, bis sie einen heftigen Erdrutsch erlebten. Martin durchsucht die Villa auf Schäden. Er wird danach traurige Klarheit haben." | "literal" |

| “item number” | “Context” | “context bias” |
|---|---|---|
| 8 | “Martin und seine Freundin haben sich vor kurzem verlobt. Sie waren überglücklich, bis sie einen heftigen Streit hatten. Martin durchdenkt die Auseinandersetzung genauer. Er wird danach traurige Klarheit haben.” | “met.” |
| 8 | “Martin und seine Freundin haben zusammen eine Villa gekauft. Sie waren überglücklich, bis sie einen heftigen Erdrutsch erlebten. Martin durchsucht die Villa auf Schäden. Er wird danach traurige Klarheit haben.” | “literal” |
| 8 | “Martin und seine Freundin haben sich vor kurzem verlobt. Sie waren überglücklich, bis sie einen heftigen Streit hatten. Martin durchdenkt die Auseinandersetzung genauer. Er wird danach traurige Klarheit haben.” | “met.” |
| 9 | “Nicole ist Kriminalbeamtin und muss eines Tages in den Slum fahren, weil es ein Gemetzel gibt. Leider kommt es zu einem heftigen Streit zwischen Schlägern aus unterschiedlichen Familien, und ein großer, aggressiver Raufbold wirkt dabei besonders gefährlich. Für Nicole wird es schwierig sein, die Situation unter Kontrolle zu bringen.” | “met.” |
| 9 | “Nicole ist Tierschützerin und muss eines Tages in den Urwald fahren, weil es ein Gemetzel gibt. Leider kommt es zu einem heftigen Streit zwischen Primaten aus unterschiedlichen Gruppen, und ein großes, aggressives Männchen wirkt dabei besonders beängstigend. Für Nicole wird es schwierig sein, die Situation unter Kontrolle zu bringen.” | “literal” |
| 9 | “Nicole ist Kriminalbeamtin und muss eines Tages in den Slum fahren, weil es ein Gemetzel gibt. Leider kommt es zu einem heftigen Streit zwischen Schlägern aus unterschiedlichen Familien, und ein großer, aggressiver Raufbold wirkt dabei besonders gefährlich. Für Nicole wird es schwierig sein, die Situation unter Kontrolle zu bringen.” | “met.” |
| 9 | “Nicole ist Tierschützerin und muss eines Tages in den Urwald fahren, weil es ein Gemetzel gibt. Leider kommt es zu einem heftigen Streit zwischen Primaten aus unterschiedlichen Gruppen, und ein großes, aggressives Männchen wirkt dabei besonders beängstigend. Für Nicole wird es schwierig sein, die Situation unter Kontrolle zu bringen.” | “literal” |
| 10 | “Eva ist Anwältin und arbeitet bei einem kleinen Tierschutzverein. Neulich kam es vor, dass ein aggressiver Raubfisch versuchte, einen Tierpfleger in einem Gewässer zu beißen. Eva wird dem Raubfisch beistehen, obwohl der Tierpfleger schwer verletzt ist. Dabei wird sie ein schlechtes Gefühl haben.” | “literal” |

| "item number" | "Context" | "context bias" |
|---|---|---|
| 10 | "Eva ist Anwältin und arbeitet bei einem großen Immobilienverwalter. Neulich kam es vor, dass ein aggressiver Hausbesitzer versuchte, eine arme Familie aus einer Wohnung zu vertreiben. Eva wird dem Hausbesitzer beistehen, obwohl die Familie in Not ist. Dabei wird sie ein schlechtes Gefühl haben." | "met." |
| 10 | "Eva ist Anwältin und arbeitet bei einem kleinen Tierschutzverein. Neulich kam es vor, dass ein aggressiver Raubfisch versuchte, einen Tierpfleger in einem Gewässer zu beißen. Eva wird dem Raubfisch beistehen, obwohl der Tierpfleger schwer verletzt ist. Dabei wird sie ein schlechtes Gefühl haben." | "literal" |
| 10 | "Eva ist Anwältin und arbeitet bei einem großen Immobilienverwalter. Neulich kam es vor, dass ein aggressiver Hausbesitzer versuchte, eine arme Familie aus einer Wohnung zu vertreiben. Eva wird dem Hausbesitzer beistehen, obwohl die Familie in Not ist. Dabei wird sie ein schlechtes Gefühl haben." | "met." |
| 11 | "Tobias ist im Urlaub auf einer kleinen Insel im Pazifik. Auf dem Weg zum Flughafen bleibt er lange im Stau stehen, weil eine alte Greisin weit vor ihm unglaublich langsam über die Straße läuft. Wenn er sie sehen wird, wird er sie durch das Fenster anschreien. Er hat Angst, ihretwegen den Flug zu verpassen." | "met." |
| 11 | "Tobias ist im Urlaub auf einer kleinen Insel im Pazifik. Auf dem Weg zum Flughafen bleibt er lange im Stau stehen, weil ein großes Reptil weit vor ihm unglaublich langsam über die Straße läuft. Wenn er es sehen wird, wird er es durch das Fenster anschreien. Er hat Angst, seinetwegen den Flug zu verpassen." | "literal" |
| 11 | "Tobias ist im Urlaub auf einer kleinen Insel im Pazifik. Auf dem Weg zum Flughafen bleibt er lange im Stau stehen, weil eine alte Greisin weit vor ihm unglaublich langsam über die Straße läuft. Wenn er sie sehen wird, wird er sie durch das Fenster anschreien. Er hat Angst, ihretwegen den Flug zu verpassen." | "met." |
| 11 | "Tobias ist im Urlaub auf einer kleinen Insel im Pazifik. Auf dem Weg zum Flughafen bleibt er lange im Stau stehen, weil ein großes Reptil weit vor ihm unglaublich langsam über die Straße läuft. Wenn er es sehen wird, wird er es durch das Fenster anschreien. Er hat Angst, seinetwegen den Flug zu verpassen." | "literal" |
| 12 | "Astrid geht zu einer Pazifikinsel mit ihrer Familie. Dort sieht sie einen Berg, aus dem unaufhörlich und laut Gas und sehr viel Lava ausgestoßen werden. Ihre Familie will den Berg bezwingen, Astrid möchte aber nicht und wird des Spaßes halber dem Berg sagen, dass sie nachher noch viel arbeiten muss. Sie wird froh sein, die Situation vermieden zu haben." | "literal" |

| “item number” | “Context” | “context bias” |
|---|---|---|
| 12 | “Astrid geht zu einer Konferenz mit ihren Freunden. Dort trifft sie einen Kollegen, der leidenschaftlich und laut argumentiert und schnell gereizt wird. Ihre Freunde wollen mit dem Kollegen essen gehen, Astrid möchte aber nicht und wird zu dem Kollegen sagen, dass sie nachher noch viel arbeiten muss. Sie wird froh sein, die Situation vermieden zu haben.” | “met.” |
| 12 | “Astrid geht zu einer Pazifikinsel mit ihrer Familie. Dort sieht sie einen Berg, aus dem unaufhörlich und laut Gas und sehr viel Lava ausgestoßen werden. Ihre Familie will den Berg bezwingen, Astrid möchte aber nicht und wird des Spaßes halber dem Berg sagen, dass sie nachher noch viel arbeiten muss. Sie wird froh sein, die Situation vermieden zu haben.” | “literal” |
| 12 | “Astrid geht zu einer Konferenz mit ihren Freunden. Dort trifft sie einen Kollegen, der leidenschaftlich und laut argumentiert und schnell gereizt wird. Ihre Freunde wollen mit dem Kollegen essen gehen, Astrid möchte aber nicht und wird zu dem Kollegen sagen, dass sie nachher noch viel arbeiten muss. Sie wird froh sein, die Situation vermieden zu haben.” | “met.” |
| 13 | “Maria und ihr Freund gehen zusammen zum Psychologen. Sie hat nämlich das Gefühl, sein wahres Ich nicht wirklich zu kennen, da er sich immer an den anderen anpasst und seine Meinung ständig ändert. Auch wenn er so mysteriös wirkt, kann sie nicht ohne ihren Partner leben. Allein bei seiner Nähe lächelt sie entspannt.” | “met.” |
| 13 | “Maria und ihr Freund gehen zusammen zum Zoo. Sie hat nämlich das Glück, eine Ausstellung gratis zu besuchen, da sie sich früher um die Tiere gekümmert hat und die Reptilien gut kennt. Auch wenn es so scheu ist, kann sie nicht ohne ihr farbenfrohes Lieblingstier leben. Allein bei seiner Nähe lächelt sie entspannt.” | “literal” |
| 13 | “Maria und ihr Freund gehen zusammen zum Psychologen. Sie hat nämlich das Gefühl, sein wahres Ich nicht wirklich zu kennen, da er sich immer an den anderen anpasst und seine Meinung ständig ändert. Auch wenn er so mysteriös wirkt, kann sie nicht ohne ihren Partner leben. Allein bei seiner Nähe lächelt sie entspannt.” | “met.” |
| 13 | “Maria und ihr Freund gehen zusammen zum Zoo. Sie hat nämlich das Glück, eine Ausstellung gratis zu besuchen, da sie sich früher um die Tiere gekümmert hat und die Reptilien gut kennt. Auch wenn es so scheu ist, kann sie nicht ohne ihr farbenfrohes Lieblingstier leben. Allein bei seiner Nähe lächelt sie entspannt.” | “literal” |

| "item number" | "Context" | "context bias" |
|---|---|---|
| 14 | "Florian ist ein Zirkus-Künstler, der gerne mit dem ungewöhnlichsten Gegner ringt. Sein anstehender Gegner ist ein bekanntes Raubtier. Florian lässt sich deswegen einen ungewöhnlichen Trick einfallen, welcher einen Knock-out für das Raubtier bedeuten wird. Er wird das Publikum damit begeistern." | "literal" |
| 14 | "Florian ist ein Faustkämpfer, der gerne den tapfersten Gegner hat. Sein anstehender Gegner ist ein sehr berühmter, edelmütiger Boxer. Florian lässt sich deswegen einen ungewöhnlichen Schlag einfallen, welcher einen Knock-out für den Boxer bedeuten wird. Er wird das Publikum damit begeistern." | "met." |
| 14 | "Florian ist ein Zirkus-Künstler, der gerne mit dem ungewöhnlichsten Gegner ringt. Sein anstehender Gegner ist ein bekanntes Raubtier. Florian lässt sich deswegen einen ungewöhnlichen Trick einfallen, welcher einen Knock-out für das Raubtier bedeuten wird. Er wird das Publikum damit begeistern." | "literal" |
| 14 | "Florian ist ein Faustkämpfer, der gerne den tapfersten Gegner hat. Sein anstehender Gegner ist ein sehr berühmter, edelmütiger Boxer. Florian lässt sich deswegen einen ungewöhnlichen Schlag einfallen, welcher einen Knock-out für den Boxer bedeuten wird. Er wird das Publikum damit begeistern." | "met." |
| 15 | "Beate will Lehrerin werden und macht deswegen ein Praktikum bei einem Erzieher. Sie wird ihm bei der Arbeit in der Schule helfen, und wird schnell merken, dass dort alles sehr chaotisch ist. Die Kinder sind laut, der Ablauf ist unstrukturiert und die Erwachsenen komplett überfordert. Es werden harte Tage für sie sein." | "met." |
| 15 | "Beate will Entertainerin werden und macht deswegen ein Praktikum bei einem Clown. Sie wird ihm bei der Arbeit in der Manege helfen, und wird schnell merken, dass dort alles sehr kompliziert ist. Die Tiere sind laut, die Jongleure sind nicht nett und das Publikum komplett gelangweilt. Es werden harte Tage für sie sein." | "literal" |
| 15 | "Beate will Lehrerin werden und macht deswegen ein Praktikum bei einem Erzieher. Sie wird ihm bei der Arbeit in der Schule helfen, und wird schnell merken, dass dort alles sehr chaotisch ist. Die Kinder sind laut, der Ablauf ist unstrukturiert und die Erwachsenen komplett überfordert. Es werden harte Tage für sie sein." | "met." |
| 15 | "Beate will Entertainerin werden und macht deswegen ein Praktikum bei einem Clown. Sie wird ihm bei der Arbeit in der Manege helfen, und wird schnell merken, dass dort alles sehr kompliziert ist. Die Tiere sind laut, die Jongleure sind nicht nett und das Publikum komplett gelangweilt. Es werden harte Tage für sie sein." | "literal" |

| “item number” | “Context” | “context bias” |
|---|---|---|
| 16 | “Charlotte zieht in ein neues Haus ein, in dem ihr Onkel gestorben ist und wo er angeblich seitdem spukt. Sie sucht nach dem Onkel, der aber unsichtbar ist und sich nur in der Nacht blicken lässt. Sie entscheidet sich, mit einem Rosenkranz durch den Keller zu gehen, doch das Fehlen von Hinweisen frustriert sie. Sie wird schnell die Geduld verlieren.” | “literal” |
| 16 | “Charlotte zieht in eine neue Gegend ein, in der ein Bär die Leute erschreckt und die Tomaten der Nachbarn frisst. Sie sucht nach dem Bären, der aber vorsichtig ist und sich äußerst selten blicken lässt. Sie entscheidet sich, mit einem Gewehr durch den Wald zu gehen, doch das Fehlen von Hinweisen frustriert sie. Sie wird schnell die Geduld verlieren.” | “met.” |
| 16 | “Charlotte zieht in ein neues Haus ein, in dem ihr Onkel gestorben ist und wo er angeblich seitdem spukt. Sie sucht nach dem Onkel, der aber unsichtbar ist und sich nur in der Nacht blicken lässt. Sie entscheidet sich, mit einem Rosenkranz durch den Keller zu gehen, doch das Fehlen von Hinweisen frustriert sie. Sie wird schnell die Geduld verlieren.” | “literal” |
| 16 | “Charlotte zieht in eine neue Gegend ein, in der ein Bär die Leute erschreckt und die Tomaten der Nachbarn frisst. Sie sucht nach dem Bären, der aber vorsichtig ist und sich äußerst selten blicken lässt. Sie entscheidet sich, mit einem Gewehr durch den Wald zu gehen, doch das Fehlen von Hinweisen frustriert sie. Sie wird schnell die Geduld verlieren.” | “met.” |
| 17 | “Katharina ist Archäologin und besucht gerade ein Stammesfest mitten in der Wüste. Sie hat einen sehr empfindlichen Magen und alle Gerichte, die ihr beim Essen angeboten werden, sind fettig, hart und überwürzt. Sie will nicht unhöflich sein, also bereitet sie sich vor, eine ordentliche Portion Fleisch zu sich zu nehmen. Ihre Entscheidung wird sie danach mehrmals bereuen.” | “met.” |
| 17 | “Katharina ist Archäologin und besucht gerade eine Ausgrabung mitten in der Wüste. Sie hat einen defekten Kompass, und alle Versuche, zum Zelt zurückzufinden, sind verzweifelt, panisch und erfolglos. Sie will nicht vor Hunger sterben und verliert den Verstand, also bereitet sie sich vor, einen großen Felsbrocken zu essen. Ihre Entscheidung wird sie danach mehrmals bereuen.” | “literal” |
| 17 | “Katharina ist Archäologin und besucht gerade ein Stammesfest mitten in der Wüste. Sie hat einen sehr empfindlichen Magen und alle Gerichte, die ihr beim Essen angeboten werden, sind fettig, hart und überwürzt. Sie will nicht unhöflich sein, also bereitet sie sich vor, eine ordentliche Portion Fleisch zu sich zu nehmen. Ihre Entscheidung wird sie danach mehrmals bereuen.” | “met.” |

| “item number” | “Context” | “context bias” |
|---|---|---|
| 17 | “Katharina ist Archäologin und besucht gerade eine Ausgrabung mitten in der Wüste. Sie hat einen defekten Kompass, und alle Versuche, zum Zelt zurückzufinden, sind verzweifelt, panisch und erfolglos. Sie will nicht vor Hunger sterben und verliert den Verstand, also bereitet sie sich vor, einen großen Felsbrocken zu essen. Ihre Entscheidung wird sie danach mehrmals bereuen.” | “literal” |
| 18 | “Daniel geht mit seinen Kindern zum Tropenhaus. Die Gehege sind voll mit wunderschönen fliegenden Insekten. Eines sitzt besonders farbenfroh und flatternd auf einer Blume. Er wird deswegen lange staunen.” | “literal” |
| 18 | “Daniel geht mit seinen Kindern ins Theater. Das Stück ist besetzt mit wunderschönen hüpfenden Tänzerinnen. Eine davon tanzt besonders elegant und leichtfüßig auf der Bühne. Er wird deswegen lange staunen.” | “met.” |
| 18 | “Daniel geht mit seinen Kindern zum Tropenhaus. Die Gehege sind voll mit wunderschönen fliegenden Insekten. Eines sitzt besonders farbenfroh und flatternd auf einer Blume. Er wird deswegen lange staunen.” | “literal” |
| 18 | “Daniel geht mit seinen Kindern ins Theater. Das Stück ist besetzt mit wunderschönen hüpfenden Tänzerinnen. Eine davon tanzt besonders elegant und leichtfüßig auf der Bühne. Er wird deswegen lange staunen.” | “met.” |
| 19 | “Robert ist Pop-Musiker und arbeitet in einer großen Plattenfirma. Obwohl die Produzenten viel Vertrauen in sein Können gesteckt haben, war Robert bis jetzt wenig erfolgreich. Aber nach langen Nächten im Studio gelingt es ihm, einen sehr eingängigen Song zu schreiben, der in kürzer Zeit überall zu hören sein wird. Er wird seinen verdienten Erfolg genießen.” | “met.” |
| 19 | “Robert ist Biologe und arbeitet in einer wissenschaftlichen Einrichtung. Obwohl die Regierung viel Geld in sein Projekt gesteckt hat, war Robert bis jetzt wenig erfolgreich. Aber nach langen Nächten im Labor gelingt es ihm, einen kritischen Erreger zu synthetisieren, der in kürzester Zeit viele Leben retten könnte. Er wird seinen verdienten Erfolg genießen.” | “literal” |
| 19 | “Robert ist Pop-Musiker und arbeitet in einer großen Plattenfirma. Obwohl die Produzenten viel Vertrauen in sein Können gesteckt haben, war Robert bis jetzt wenig erfolgreich. Aber nach langen Nächten im Studio gelingt es ihm, einen sehr eingängigen Song zu schreiben, der in kürzer Zeit überall zu hören sein wird. Er wird seinen verdienten Erfolg genießen.” | “met.” |

| “item number” | “Context” | “context bias” |
|---|---|---|
| 19 | “Robert ist Biologe und arbeitet in einer wissenschaftlichen Einrichtung. Obwohl die Regierung viel Geld in sein Projekt gesteckt hat, war Robert bis jetzt wenig erfolgreich. Aber nach langen Nächten im Labor gelingt es ihm, einen kritischen Erreger zu synthetisieren, der in kürzester Zeit viele Leben retten könnte. Er wird seinen verdienten Erfolg genießen.” | “literal” |
| 20 | “Stephanie ist Tiermedizinerin und meldet sich bei einem Pflegedienst an. Sie kümmert sich im Zoo um eine verletzte, schüchterne atlantische Robbe und füttert sie regelmäßig. Sie ist sehr überrascht zu sehen, dass die Robbe unglaublich schlecht gelaunt ist, und nichts fressen will. Nichtdestotrotz versucht sie, eine gute Zeit mit ihr zu verbringen.” | “literal” |
| 20 | “Stephanie ist Single und meldet sich bei einer Dating-Website an. Sie flirtet im Internet mit einem netten, angeblich gutaussehenden Mann und lädt ihn zum Essen ein. Sie ist sehr überrascht zu sehen, dass der Mann unglaublich fettleibig ist, und kaum durch die Wohnungstür passt. Nichtdestotrotz versucht sie, eine gute Zeit mit ihm zu verbringen.” | “met.” |
| 20 | “Stephanie ist Tiermedizinerin und meldet sich bei einem Pflegedienst an. Sie kümmert sich im Zoo um eine verletzte, schüchterne atlantische Robbe und füttert sie regelmäßig. Sie ist sehr überrascht zu sehen, dass die Robbe unglaublich schlecht gelaunt ist, und nichts fressen will. Nichtdestotrotz versucht sie, eine gute Zeit mit ihr zu verbringen.” | “literal” |
| 20 | “Stephanie ist Single und meldet sich bei einer Dating-Website an. Sie flirtet im Internet mit einem netten, angeblich gutaussehenden Mann und lädt ihn zum Essen ein. Sie ist sehr überrascht zu sehen, dass der Mann unglaublich fettleibig ist, und kaum durch die Wohnungstür passt. Nichtdestotrotz versucht sie, eine gute Zeit mit ihm zu verbringen.” | “met.” |
| 21 | “Christian geht am Nachmittag in die Schule um seine Tochter abzuholen, da es ihr nicht gut geht. Sie ist sehr zerbrechlich und emotional, und sie wird anfangen zu weinen, sobald Christian sie anfasst. Er wird aber trotzdem seine Arme ausstrecken und die Zeit mit der sensiblen Tochter genießen. Es wird danach ein toller Tag werden.” | “met.” |
| 21 | “Christian geht im Winter in den Park um die Eiskristalle zu sehen, die vom Himmel fallen. Sie sind sehr symmetrisch und einzigartig, und sie werden anfangen zu schmelzen, sobald Christian sie anfasst. Er wird aber trotzdem seine Arme ausstrecken und den tollen Anblick der Eiskristalle genießen. Es wird danach ein toller Tag werden.” | “literal” |

| “item number” | “Context” | “context bias” |
|---|---|---|
| 21 | “Christian geht am Nachmittag in die Schule um seine Tochter abzuholen, da es ihr nicht gut geht. Sie ist sehr zerbrechlich und emotional, und sie wird anfangen zu weinen, sobald Christian sie anfasst. Er wird aber trotzdem seine Arme ausstrecken und die Zeit mit der sensiblen Tochter genießen. Es wird danach ein toller Tag werden.” | “met.” |
| 21 | “Christian geht im Winter in den Park um die Eiskristalle zu sehen, die vom Himmel fallen. Sie sind sehr symmetrisch und einzigartig, und sie werden anfangen zu schmelzen, sobald Christian sie anfasst. Er wird aber trotzdem seine Arme ausstrecken und den tollen Anblick der Eiskristalle genießen. Es wird danach ein toller Tag werden.” | “literal” |
| 22 | “Petra ist Ermittlerin und ist mit einem schwierigen Fall beschäftigt. Neulich entdeckte sie einen Holzsplitter am Tatort und wusste nicht, ob er für den Fall entscheidend ist. Er ist klein, angebrochen und blutverschmiert. Sie wird das Ganze nicht ernstnehmen.” | “literal” |
| 22 | “Petra ist Trainerin einer Rugby Mannschaft und freut sich über neue Zugänge. Neulich entdeckte sie einen Teenager und wusste nicht, ob er ins Team passt. Er ist sehr dünn, zerbrechlich und den anderen körperlich unterlegen. Sie wird das Ganze nicht ernstnehmen.” | “met.” |
| 22 | “Petra ist Ermittlerin und ist mit einem schwierigen Fall beschäftigt. Neulich entdeckte sie einen Holzsplitter am Tatort und wusste nicht, ob er für den Fall entscheidend ist. Er ist klein, angebrochen und blutverschmiert. Sie wird das Ganze nicht ernstnehmen.” | “literal” |
| 22 | “Petra ist Trainerin einer Rugby Mannschaft und freut sich über neue Zugänge. Neulich entdeckte sie einen Teenager und wusste nicht, ob er ins Team passt. Er ist sehr dünn, zerbrechlich und den anderen körperlich unterlegen. Sie wird das Ganze nicht ernstnehmen.” | “met.” |
| 23 | “Lukas ist verbeamtet und arbeitet in einem kleinen Team. Eines Tages hat ein Feuerwehrmann, der nie auf die anderen hört, eine riskante Entscheidung getroffen, obwohl Lukas ihm davon abgeraten hat. Es hat sich aber gezeigt, dass der Feuerwehrmann Recht hatte, und Lukas wird ihm dafür gratulieren müssen. Die Beziehung zwischen den beiden ist nicht leicht.” | “met.” |
| 23 | “Lukas ist Bauer und arbeitet in einem kleinen Bauernhof. Eines Tages hat ein Nutztier, das nie auf Lukas hört, absolut nichts fressen wollen, obwohl Lukas es dazu animiert hat. Es hat sich aber gezeigt, dass das Nutztier gespürt hat, dass das Futter faulig war, und Lukas wird ihm dafür gratulieren müssen. Die Beziehung zwischen den beiden ist nicht leicht.” | “literal” |

| "item number" | "Context" | "context bias" |
|---|---|---|
| 23 | "Lukas ist verbeamtet und arbeitet in einem kleinen Team. Eines Tages hat ein Feuerwehrmann, der nie auf die anderen hört, eine riskante Entscheidung getroffen, obwohl Lukas ihm davon abgeraten hat. Es hat sich aber gezeigt, dass der Feuerwehrmann Recht hatte, und Lukas wird ihm dafür gratulieren müssen. Die Beziehung zwischen den beiden ist nicht leicht." | "met." |
| 23 | "Lukas ist Bauer und arbeitet in einem kleinen Bauernhof. Eines Tages hat ein Nutztier, das nie auf Lukas hört, absolut nichts fressen wollen, obwohl Lukas es dazu animiert hat. Es hat sich aber gezeigt, dass das Nutztier gespürt hat, dass das Futter faulig war, und Lukas wird ihm dafür gratulieren müssen. Die Beziehung zwischen den beiden ist nicht leicht." | "literal" |
| 24 | "Jonas interessiert sich sehr für Design und entscheidet sich, zu einer Messe zu gehen. Er sieht ein Möbelstück, das fast doppelt so hoch ist wie alle anderen. Er wird sehr begeistert von ihm sein und wird glauben, dass das Möbelstück den Wettbewerb gewinnen wird. In wenigen Minuten wird es tatsächlich passieren." | "literal" |
| 24 | "Jonas interessiert sich sehr für Basketball und entscheidet sich, ins Stadion zu gehen. Er sieht ein Spieler, der fast doppelt so groß und schwer ist wie die anderen. Er wird sehr begeistert von ihm sein und wird wetten, dass der Spieler das Spiel gewinnen wird. In wenigen Minuten wird es tatsächlich passieren." | "met." |
| 24 | "Jonas interessiert sich sehr für Design und entscheidet sich, zu einer Messe zu gehen. Er sieht ein Möbelstück, das fast doppelt so hoch ist wie alle anderen. Er wird sehr begeistert von ihm sein und wird glauben, dass das Möbelstück den Wettbewerb gewinnen wird. In wenigen Minuten wird es tatsächlich passieren." | "literal" |
| 24 | "Jonas interessiert sich sehr für Basketball und entscheidet sich, ins Stadion zu gehen. Er sieht ein Spieler, der fast doppelt so groß und schwer ist wie die anderen. Er wird sehr begeistert von ihm sein und wird wetten, dass der Spieler das Spiel gewinnen wird. In wenigen Minuten wird es tatsächlich passieren." | "met." |
| 25 | "Emily ist Studentin und besucht eine Vorlesung. Der Professor der Vorlesung ist ein Vertreter von veralteten Theorien, und wird von den jungen Generationen nicht mehr ernst genommen. Emily aber meldet sich ständig im Kurs, und stellt die Weltsicht des Professors in Frage. Dies wird ihr übelgenommen werden." | "met." |
| 25 | "Emily ist Studentin und besucht ein Museum. Die Versteinerungen des Museums sind Überreste vergangener Zeiten, und werden von jungen Generationen nicht mehr geschätzt. Emily aber schießt Bilder von einer Versteinerung, und setzt sich über das Fotoverbot des Museums hinweg. Dies wird ihr übelgenommen werden." | "literal" |

| "item number" | "Context" | "context bias" |
|---|---|---|
| 25 | "Emily ist Studentin und besucht eine Vorlesung. Der Professor der Vorlesung ist ein Vertreter von veralteten Theorien, und wird von den jungen Generationen nicht mehr ernst genommen. Emily aber meldet sich ständig im Kurs, und stellt die Weltsicht des Professors in Frage. Dies wird ihr übelgenommen werden." | "met." |
| 25 | "Emily ist Studentin und besucht ein Museum. Die Versteinerungen des Museums sind Überreste vergangener Zeiten, und werden von jungen Generationen nicht mehr geschätzt. Emily aber schießt Bilder von einer Versteinerung, und setzt sich über das Fotoverbot des Museums hinweg. Dies wird ihr übelgenommen werden." | "literal" |
| 26 | "David geht im Winter in eine Ausstellung im Freiem, um sich Skulpturen aus Schnee anzusehen. Er ist sehr betrunken und merkt nicht, dass die Skulptur, die er für einen Menschen hält, ein Teil der Sammlung ist. Nach einer Weile wird er es merken und sich schämen." | "literal" |
| 26 | "David geht im Winter in eine Bar in der Gegend, um Frauen kennenzulernen. Er ist sehr betrunken und merkt nicht, dass die Blondine, die an der Theke alleine trinkt, nicht reden will und ihn gefühllos ignoriert. Nach einer Weile wird er es merken und sich schämen." | "met." |
| 26 | "David geht im Winter in eine Ausstellung im Freiem, um sich Skulpturen aus Schnee anzusehen. Er ist sehr betrunken und merkt nicht, dass die Skulptur, die er für einen Menschen hält, ein Teil der Sammlung ist. Nach einer Weile wird er es merken und sich schämen." | "literal" |
| 26 | "David geht im Winter in eine Bar in der Gegend, um Frauen kennenzulernen. Er ist sehr betrunken und merkt nicht, dass die Blondine, die an der Theke alleine trinkt, nicht reden will und ihn gefühllos ignoriert. Nach einer Weile wird er es merken und sich schämen." | "met." |
| 27 | "Vanessa arbeitet für eine Wohnungsgesellschaft und muss demnächst eine freigewordene Wohnung besichtigen. Sie geht in die Wohnung hinein und merkt, dass alles chaotisch ist. Die ist zugemüllt und schwer begehbar. Sie wird eine schwierige Aufgabe haben." | "met." |
| 27 | "Vanessa arbeitet für einen Sender und muss demnächst eine bezaubernde Dokumentation über den Regenwald drehen. Sie geht in den Regenwald und merkt, dass alles chaotisch ist. Der ist sehr dunkel und schwer begehbar. Sie wird eine schwierige Aufgabe haben." | "literal" |

| **"item number"** | **"Context"** | **"context bias"** |
|---|---|---|
| 27 | "Vanessa arbeitet für eine Wohnungsgesellschaft und muss demnächst eine freigewordene Wohnung besichtigen. Sie geht in die Wohnung hinein und merkt, dass alles chaotisch ist. Die ist zugemüllt und schwer begehbar. Sie wird eine schwierige Aufgabe haben." | "met." |
| 27 | "Vanessa arbeitet für einen Sender und muss demnächst eine bezaubernde Dokumentation über den Regenwald drehen. Sie geht in den Regenwald und merkt, dass alles chaotisch ist. Der ist sehr dunkel und schwer begehbar. Sie wird eine schwierige Aufgabe haben." | "literal" |
| 28 | "Karoline arbeitet in einem Krankenhaus und freut sich immer darauf, Patienten in Not helfen zu können. Eines Tages kommt ein alter Mann, der eine offene Wunde hat, zu ihr in die Notaufnahme, und will einen Wundverband bekommen. Karoline weiß, dass die Binde ihm kaum nützen wird, will ihm aber trotzdem helfen. Die Hilfe wird nur eine Weile anhalten." | "literal" |
| 28 | "Karoline arbeitet in einem Unternehmen und freut sich immer darauf, ihren Angestellten helfen zu können. Eines Tages kommt ihr Assistent, der eine starke Depression hat, zu ihr ins Büro, und will eine Arbeitsauszeit bekommen. Karoline weiß, dass der Urlaub ihm kaum nützen wird, will ihm aber trotzdem helfen. Die Hilfe wird nur eine Weile anhalten." | "met." |
| 28 | "Karoline arbeitet in einem Krankenhaus und freut sich immer darauf, Patienten in Not helfen zu können. Eines Tages kommt ein alter Mann, der eine offene Wunde hat, zu ihr in die Notaufnahme, und will einen Wundverband bekommen. Karoline weiß, dass die Binde ihm kaum nützen wird, will ihm aber trotzdem helfen. Die Hilfe wird nur eine Weile anhalten." | "literal" |
| 28 | "Karoline arbeitet in einem Unternehmen und freut sich immer darauf, ihren Angestellten helfen zu können. Eines Tages kommt ihr Assistent, der eine starke Depression hat, zu ihr ins Büro, und will eine Arbeitsauszeit bekommen. Karoline weiß, dass der Urlaub ihm kaum nützen wird, will ihm aber trotzdem helfen. Die Hilfe wird nur eine Weile anhalten." | "met." |
| 29 | "Laura arbeitet als Sportlehrerin in einem Gymnasium. Einer ihrer Schüler ist besonders begabt in der Leichtathletik. Er ist unglaublich schnell und gewinnt mühelos alle Wettbewerbe. Laura will ihm helfen, seine Zeiten weiter zu verbessern. Die Mühe wird sich auszahlen." | "met." |

| **"item number"** | **"Context"** | **"context bias"** |
|---|---|---|
| 29 | "Laura arbeitet als Tierpflegerin in einem Tierpark. Eine der Raubkatzen ist besonders aggressiv zu den Pflegern. Sie ist unglaublich reizbar und erschreckt ständig alle Parkbesucher. Laura will ihr helfen, ihr Benehmen weiter zu verbessern. Die Mühe wird sich auszahlen." | "literal" |
| 29 | "Laura arbeitet als Sportlehrerin in einem Gymnasium. Einer ihrer Schüler ist besonders begabt in der Leichtathletik. Er ist unglaublich schnell und gewinnt mühelos alle Wettbewerbe. Laura will ihm helfen, seine Zeiten weiter zu verbessern. Die Mühe wird sich auszahlen." | "met." |
| 29 | "Laura arbeitet als Tierpflegerin in einem Tierpark. Eine der Raubkatzen ist besonders aggressiv zu den Pflegern. Sie ist unglaublich reizbar und erschreckt ständig alle Parkbesucher. Laura will ihr helfen, ihr Benehmen weiter zu verbessern. Die Mühe wird sich auszahlen." | "literal" |
| 30 | "Philipp ist neu im Dorf und will eine Freizeitbeschäftigung haben. Er fängt an, Zeit mit einem Kriechtier zu verbringen, das den Ruf hat, gefühllos und tödlich zu sein. Ihn interessiert es nicht und entwickelt langsam Vertrauen dem Tier gegenüber. Er wird es nicht einschätzen können, wie groß sein Fehler ist." | "literal" |
| 30 | "Philipp ist neu im Dorf und will nette Leute kennenlernen. Er fängt an, Zeit mit einer Hausfrau zu verbringen, die den Ruf hat, hinterlistig und heimtückisch zu sein. Ihn interessiert das nicht und er entwickelt großes Vertrauen der Frau gegenüber. Er wird nicht einschätzen können, wie groß sein Fehler ist." | "met." |
| 30 | "Philipp ist neu im Dorf und will eine Freizeitbeschäftigung haben. Er fängt an, Zeit mit einem Kriechtier zu verbringen, das den Ruf hat, gefühllos und tödlich zu sein. Ihn interessiert es nicht und entwickelt langsam Vertrauen dem Tier gegenüber. Er wird es nicht einschätzen können, wie groß sein Fehler ist." | "literal" |
| 30 | "Philipp ist neu im Dorf und will nette Leute kennenlernen. Er fängt an, Zeit mit einer Hausfrau zu verbringen, die den Ruf hat, hinterlistig und heimtückisch zu sein. Ihn interessiert das nicht und er entwickelt großes Vertrauen der Frau gegenüber. Er wird nicht einschätzen können, wie groß sein Fehler ist." | "met." |
| 31 | "Nadine ist Angestellte und mag ihren Vorgesetzten nicht. Er ist ein großer, wütender Mann, der äußerst aufdringlich ist und seinen Willen über alle anderen gnadenlos durchsetzt. Eines Tages wird es Nadine nicht mehr aushalten und ihren Vorgesetzten anschreien, weil er sie zu sehr nötigen wird. Es wird zu einer tödlichen Stille kommen." | "met." |

| “item number” | “Context” | “context bias” |
|---|---|---|
| 31 | “Nadine ist Soldatin und mag ihren Kampfwagen nicht. Der ist eine große, veraltete Maschine, die äußerst langsam ist und selten verlässlich funktioniert. Eines Tages wird es Nadine nicht mehr aushalten und ihren Kampfwagen wiederholt anschreien, weil er sich nicht mehr fortbewegen wird. Es wird danach zu einer tödlichen Stille kommen.” | “literal” |
| 31 | “Nadine ist Angestellte und mag ihren Vorgesetzten nicht. Er ist ein großer, wütender Mann, der äußerst aufdringlich ist und seinen Willen über alle anderen gnadenlos durchsetzt. Eines Tages wird es Nadine nicht mehr aushalten und ihren Vorgesetzten anschreien, weil er sie zu sehr nötigen wird. Es wird zu einer tödlichen Stille kommen.” | “met.” |
| 31 | “Nadine ist Soldatin und mag ihren Kampfwagen nicht. Der ist eine große, veraltete Maschine, die äußerst langsam ist und selten verlässlich funktioniert. Eines Tages wird es Nadine nicht mehr aushalten und ihren Kampfwagen wiederholt anschreien, weil er sich nicht mehr fortbewegen wird. Es wird danach zu einer tödlichen Stille kommen.” | “literal” |
| 32 | “Sebastian liebt eine berühmte Adlige. Er hat sie in einem Schloss kennengelernt und seitdem sind sie unzertrennlich. Die Adlige ist schwach und abhängig, und kann sehr hilfsbedürftig sein. Deswegen tut Sebastian alles für sie, wenn sie etwas braucht. Er wird sich immer um sie kümmern wollen.” | “literal” |
| 32 | “Sebastian liebt eine wunderschöne Katze. Er hat sie in einem Tierheim adoptiert und seitdem sind sie unzertrennlich. Die Katze ist verwöhnt und launisch, und kann sehr wählerisch sein. Deswegen würde Sebastian alles für sie tun, wenn sie etwas braucht. Er wird sich immer um sie kümmern wollen.” | “met.” |
| 32 | “Sebastian liebt eine berühmte Adlige. Er hat sie in einem Schloss kennengelernt und seitdem sind sie unzertrennlich. Die Adlige ist schwach und abhängig, und kann sehr hilfsbedürftig sein. Deswegen tut Sebastian alles für sie, wenn sie etwas braucht. Er wird sich immer um sie kümmern wollen.” | “literal” |
| 32 | “Sebastian liebt eine wunderschöne Katze. Er hat sie in einem Tierheim adoptiert und seitdem sind sie unzertrennlich. Die Katze ist verwöhnt und launisch, und kann sehr wählerisch sein. Deswegen würde Sebastian alles für sie tun, wenn sie etwas braucht. Er wird sich immer um sie kümmern wollen.” | “met.” |

| “item number” | “Context” | “context bias” |
|---|---|---|
| 33 | “Fabian arbeitet in einer Kita und begrüßt die neu hinzugekommenen Kinder. Ein Junge ist sehr laut und redselig und kann einfach nicht aufhören, die anderen Kinder nachzuahmen. Der Junge scheint sehr große Freude am Reden zu haben. Fabian wird sich für diesen Fall besonders viel Zeit nehmen.” | “met.” |
| 33 | “Fabian arbeitet in einem Tierheim und begrüßt die neu hinzugekommenen Vögel. Einer ist sehr laut und redselig und kann einfach nicht aufhören, die anderen Vögel nachzuahmen. Der Vogel scheint sehr große Freude am Streiten zu haben. Fabian wird sich für diesen Fall besonders viel Zeit nehmen.” | “literal” |
| 33 | “Fabian arbeitet in einer Kita und begrüßt die neu hinzugekommenen Kinder. Ein Junge ist sehr laut und redselig und kann einfach nicht aufhören, die anderen Kinder nachzuahmen. Der Junge scheint sehr große Freude am Reden zu haben. Fabian wird sich für diesen Fall besonders viel Zeit nehmen.” | “met.” |
| 33 | “Fabian arbeitet in einem Tierheim und begrüßt die neu hinzugekommenen Vögel. Einer ist sehr laut und redselig und kann einfach nicht aufhören, die anderen Vögel nachzuahmen. Der Vogel scheint sehr große Freude am Streiten zu haben. Fabian wird sich für diesen Fall besonders viel Zeit nehmen.” | “literal” |
| 34 | “Alexander ist Touristenführer und leitet gerade einen Ausflug auf einem Boot am Amazonas. Einer der Fische im Fluss wirkt plötzlich sehr hungrig und aggressiv, und versucht, alle anderen Fische einzuschüchtern. Alexander wird zweifelsohne einschreiten.” | “literal” |
| 34 | “Alexander ist Unternehmensberater und leitet gerade einen Workshop über Kommunikation am Arbeitsplatz. Einer der Teilnehmer wirkt arrogant und egoistisch, und versucht, alle anderen Teilnehmer einzuschüchtern. Alexander wird zweifelsohne einschreiten.” | “met.” |
| 34 | “Alexander ist Touristenführer und leitet gerade einen Ausflug auf einem Boot am Amazonas. Einer der Fische im Fluss wirkt plötzlich sehr hungrig und aggressiv, und versucht, alle anderen Fische einzuschüchtern. Alexander wird zweifelsohne einschreiten.” | “literal” |
| 34 | “Alexander ist Unternehmensberater und leitet gerade einen Workshop über Kommunikation am Arbeitsplatz. Einer der Teilnehmer wirkt arrogant und egoistisch, und versucht, alle anderen Teilnehmer einzuschüchtern. Alexander wird zweifelsohne einschreiten.” | “met.” |

| **"item number"** | **"Context"** | **"context bias"** |
|---|---|---|
| 35 | "Simone ist eine einflussreiche Senatorin und hat sich einen politischen Ziehsohn ausgesucht. Er ist sehr klug, aber neigt dazu, habgierig zu sein und die Schwächen anderer Menschen auszunutzen. Simone wird sich deswegen um seinen Ruf bemühen müssen. Sie wird dafür hart arbeiten." | "met." |
| 35 | "Simone ist eine einflussreiche Tiermedizinerin und hat einen seltenen Greifvogel erforscht. Er ist sehr klug, aber neigt dazu, aggressiv zu sein und die Schwächen der anderen Vögel auszunutzen. Simone wird sich deswegen um seinen Ruf bemühen müssen. Sie wird dafür hart arbeiten." | "literal" |
| 35 | "Simone ist eine einflussreiche Senatorin und hat sich einen politischen Ziehsohn ausgesucht. Er ist sehr klug, aber neigt dazu, habgierig zu sein und die Schwächen anderer Menschen auszunutzen. Simone wird sich deswegen um seinen Ruf bemühen müssen. Sie wird dafür hart arbeiten." | "met." |
| 35 | "Simone ist eine einflussreiche Tiermedizinerin und hat einen seltenen Greifvogel erforscht. Er ist sehr klug, aber neigt dazu, aggressiv zu sein und die Schwächen der anderen Vögel auszunutzen. Simone wird sich deswegen um seinen Ruf bemühen müssen. Sie wird dafür hart arbeiten." | "literal" |
| 36 | "Niklas konkurriert häufig um einen sehr berühmten Sportpokal. Er will den Wettbewerb gewinnen nur um sein Ego zu stärken. Er genießt es auch sehr, damit vor seinen Freunden zu prahlen. Es wird für ihn ein harter Kampf sein." | "literal" |
| 36 | "Niklas flirtet häufig mit einer berühmten Sängerin. Er will ihre Liebe erobern nur um sein Selbstwertgefühl zu stärken. Er genießt es auch sehr, damit vor seinen Freunden zu prahlen. Es wird für ihn ein harter Kampf sein." | "met." |
| 36 | "Niklas konkurriert häufig um einen sehr berühmten Sportpokal. Er will den Wettbewerb gewinnen nur um sein Ego zu stärken. Er genießt es auch sehr, damit vor seinen Freunden zu prahlen. Es wird für ihn ein harter Kampf sein." | "literal" |
| 36 | "Niklas flirtet häufig mit einer berühmten Sängerin. Er will ihre Liebe erobern nur um sein Selbstwertgefühl zu stärken. Er genießt es auch sehr, damit vor seinen Freunden zu prahlen. Es wird für ihn ein harter Kampf sein." | "met." |

## B.2 Critical Items for Experiment 5-6

**Table B.2:** critical items of Experiments 5 and 6

| **"item number"** | **"Target sentence"** | **"Verb constraint"** | **"context bias"** |
|---|---|---|---|
| 1 | "Sarah kontert einen Bulldozer, und wird offensichtlich den Angestellten aufhalten." | "early" | "metaphoric" |
| 1 | "Sarah kontert einen Bulldozer, und wird offensichtlich das Raupenfahrzeug aufhalten." | "early" | "literal" |
| 1 | "Sarah wird einen Bulldozer kontern, und wird offensichtlich den Angestellten aufhalten." | "late" | "metaphoric" |
| 1 | "Sarah wird einen Bulldozer kontern, und wird offensichtlich das Raupenfahrzeug aufhalten." | "late" | "literal" |
| 2 | "Michael wird einen Drachen belästigen, und wird vorsichtig die Feuerechse besänftigen." | "late" | "literal" |
| 2 | "Michael belästigt einen Drachen, und wird vorsichtig die Bibliothekarin besänftigen." | "early" | "metaphoric" |
| 2 | "Michael belästigt einen Drachen, und wird vorsichtig die Feuerechse besänftigen." | "early" | "literal" |
| 2 | "Michael wird einen Drachen belästigen, und wird vorsichtig die Bibliothekarin besänftigen." | "late" | "metaphoric" |
| 3 | "Thomas wird einen Roboter liebkosen und wird hoffentlich die Oma zum Lächeln bringen." | "late" | "metaphoric" |
| 3 | "Thomas wird einen Roboter liebkosen und wird hoffentlich die Maschine zum Lächeln bringen." | "late" | "literal" |
| 3 | "Thomas liebkost einen Roboter und wird hoffentlich die Oma zum Lächeln bringen." | "early" | "metaphoric" |
| 3 | "Thomas liebkost einen Roboter und wird hoffentlich die Maschine zum Lächeln bringen." | "early" | "literal" |
| 4 | "Karin trainiert ein Altersheim, und wird erwartungsvoll die Senioren kräftigen." | "early" | "literal" |
| 4 | "Karin wird ein Altersheim trainieren, und wird erwartungsvoll die Fußballspieler kräftigen." | "late" | "metaphoric" |
| 4 | "Karin wird ein Altersheim trainieren, und wird erwartungsvoll die Senioren kräftigen." | "late" | "literal" |
| 4 | "Karin trainiert ein Altersheim, und wird erwartungsvoll die Fußballspieler kräftigen." | "early" | "metaphoric" |

| "item number" | "Target sentence" | "Verb constraint" | "context bias" |
|---|---|---|---|
| 5 | "Andreas berät einen Engel und wird wiederholt die Nonne ansprechen." | "early" | "metaphoric" |
| 5 | "Andreas berät einen Engel und wird wiederholt das Gemälde ansprechen." | "early" | "literal" |
| 5 | "Andreas wird einen Engel beraten und wird wiederholt die Nonne ansprechen." | "late" | "metaphoric" |
| 5 | "Andreas wird einen Engel beraten und wird wiederholt das Gemälde ansprechen." | "late" | "literal" |
| 6 | "Stefan wird eine Hyäne interviewen und wird unerbittlich das Raubtier porträtieren." | "late" | "literal" |
| 6 | "Stefan interviewt eine Hyäne und wird unerbittlich den Makler porträtieren." | "early" | "metaphoric" |
| 6 | "Stefan interviewt eine Hyäne und wird unerbittlich das Raubtier porträtieren." | "early" | "literal" |
| 6 | "Stefan wird eine Hyäne interviewen und wird unerbittlich den Makler porträtieren." | "late" | "metaphoric" |
| 7 | "Susanne wird einen Dinosaurier grüßen und wird lange Zeit den Politiker anlächeln." | "late" | "metaphoric" |
| 7 | "Susanne wird einen Dinosaurier grüßen und wird lange Zeit das Riesenskelett anlächeln." | "late" | "literal" |
| 7 | "Susanne grüßt einen Dinosaurier und wird lange Zeit den Politiker anlächeln." | "early" | "metaphoric" |
| 7 | "Susanne grüßt einen Dinosaurier und wird lange Zeit das Riesenskelett anlächeln." | "early" | "literal" |
| 8 | "Martin empfindet einen Riss, und wird schließlich die Villa erneuern." | "early" | "literal" |
| 8 | "Martin wird einen Riss empfinden, und wird schließlich die Verlobung erneuern." | "late" | "metaphoric" |
| 8 | "Martin wird einen Riss empfinden, und wird schließlich die Villa erneuern." | "late" | "literal" |
| 8 | "Martin empfindet einen Riss, und wird schließlich die Verlobung erneuern." | "early" | "metaphoric" |
| 9 | "Nicole verhaftet einen Gorilla, und wird sicherlich den Raufbold hart anpacken." | "early" | "metaphoric" |
| 9 | "Nicole verhaftet einen Gorilla, und wird sicherlich das Männchen hart anpacken." | "early" | "literal" |
| 9 | "Nicole wird einen Gorilla verhaften, und wird sicherlich den Raufbold hart anpacken." | "late" | "metaphoric" |
| 9 | "Nicole wird einen Gorilla verhaften, und wird sicherlich das Männchen hart anpacken." | "late" | "literal" |
| 10 | "Eva wird einen Hai vertreten, und wird augenscheinlich den Raubfisch beschützen." | "late" | "literal" |

| “item number” | “Target sentence” | “Verb constraint” | “context bias” |
|---|---|---|---|
| 10 | “Eva vertritt einen Hai, und wird augenscheinlich den Hausbesitzer beschützen.” | “early” | “metaphoric” |
| 10 | “Eva vertritt einen Hai, und wird augenscheinlich den Raubfisch beschützen.” | “early” | “literal” |
| 10 | “Eva wird einen Hai vertreten, und wird augenscheinlich den Hausbesitzer beschützen.” | “late” | “metaphoric” |
| 11 | “Tobias wird eine Schildkröte beschimpfen und wird erbittert die Greisin betrachten.” | “late” | “metaphoric” |
| 11 | “Tobias wird eine Schildkröte beschimpfen und wird erbittert das Reptil betrachten.” | “late” | “literal” |
| 11 | “Tobias beschimpft eine Schildkröte und wird erbittert die Greisin betrachten.” | “early” | “metaphoric” |
| 11 | “Tobias beschimpft eine Schildkröte und wird erbittert das Reptil betrachten.” | “early” | “literal” |
| 12 | “Astrid belügt einen Vulkan und wird erfolgreich dem Berg entkommen.” | “early” | “literal” |
| 12 | “Astrid wird einen Vulkan belügen und wird erfolgreich dem Kollegen entkommen.” | “late” | “metaphoric” |
| 12 | “Astrid wird einen Vulkan belügen und wird erfolgreich dem Berg entkommen.” | “late” | “literal” |
| 12 | “Astrid belügt einen Vulkan und wird erfolgreich dem Kollegen entkommen.” | “early” | “metaphoric” |
| 13 | “Maria küsst ein Chamäleon und wird versehentlich den Partner verängstigen.” | “early” | “metaphoric” |
| 13 | “Maria küsst ein Chamäleon, und wird versehentlich das Lieblingstier verängstigen.” | “early” | “literal” |
| 13 | “Maria wird ein Chamäleon küssen, und wird versehentlich den Partner verängstigen.” | “late” | “metaphoric” |
| 13 | “Maria wird ein Chamäleon küssen, und wird versehentlich das Lieblingstier verängstigen.” | “late” | “literal” |
| 14 | “Florian wird einen Löwen bekämpfen, und wird auf jeden Fall das Raubtier besiegen.” | “late” | “literal” |
| 14 | “Florian bekämpft einen Löwen, und wird auf jeden Fall den Boxer besiegen.” | “early” | “metaphoric” |
| 14 | “Florian bekämpft einen Löwen, und wird auf jeden Fall das Raubtier besiegen.” | “early” | “literal” |
| 14 | “Florian wird einen Löwen bekämpfen, und wird auf jeden Fall den Boxer besiegen.” | “late” | “metaphoric” |
| 15 | “Beate wird einen Zirkus umsorgen, und wird binnen kurzem die Schule verachten.” | “late” | “metaphoric” |

| “item number” | “Target sentence” | “Verb constraint” | “context bias” |
|---|---|---|---|
| 15 | “Beate wird einen Zirkus umsorgen und wird binnen kurzem die Manege verachten.” | “late” | “literal” |
| 15 | “Beate umsorgt einen Zirkus, und wird binnen kurzem die Schule verachten.” | “early” | “metaphoric” |
| 15 | “Beate umsorgt einen Zirkus, und wird binnen kurzem die Manege verachten.” | “early” | “literal” |
| 16 | “Charlotte jagt ein Gespenst, und wird hoffnungslos den Onkel fortlassen.” | “early” | “literal” |
| 16 | “Charlotte wird ein Gespenst jagen, und wird hoffnungslos den Bären fortlassen.” | “late” | “metaphoric” |
| 16 | “Charlotte wird ein Gespenst jagen, und wird hoffnungslos den Onkel fortlassen.” | “late” | “literal” |
| 16 | “Charlotte jagt ein Gespenst, und wird hoffnungslos den Bären fortlassen.” | “early” | “metaphoric” |
| 17 | “Katharina verzehrt einen Stein und wird schnellstmöglich das Fleisch erbrechen.” | “early” | “metaphoric” |
| 17 | “Katharina verzehrt einen Stein und wird schnellstmöglich den Felsbrocken erbrechen.” | “early” | “literal” |
| 17 | “Katharina wird einen Stein verzehren und wird schnellstmöglich das Fleisch erbrechen.” | “late” | “metaphoric” |
| 17 | “Katharina wird einen Stein verzehren und wird schnellstmöglich den Felsbrocken erbrechen.” | “late” | “literal” |
| 18 | “Daniel wird einen Schmetterling begehren und wird zweifelsohne das Insekt wiedersehen.” | “late” | “literal” |
| 18 | “Daniel begehrt einen Schmetterling und wird zweifelsohne die Tänzerin wiedersehen.” | “early” | “metaphoric” |
| 18 | “Daniel begehrt einen Schmetterling und wird zweifelsohne das Insekt wiedersehen.” | “early” | “literal” |
| 18 | “Daniel wird einen Schmetterling begehren und wird zweifelsohne die Tänzerin wiedersehen.” | “late” | “metaphoric” |
| 19 | “Robert wird einen Virus komponieren, und wird erfolgreich den Song vermarkten.” | “late” | “metaphoric” |
| 19 | “Robert wird einen Virus komponieren, und wird erfolgreich den Erreger vermarkten.” | “late” | “literal” |
| 19 | “Robert komponiert einen Virus, und wird erfolgreich den Song vermarkten.” | “early” | “metaphoric” |
| 19 | “Robert komponiert einen Virus, und wird erfolgreich den Erreger vermarkten.” | “early” | “literal” |

| "item number" | "Target sentence" | "Verb constraint" | "context bias" |
|---|---|---|---|
| 20 | "Stephanie bekocht ein Walross, und wird permanent die Robbe anlächeln." | "early" | "literal" |
| 20 | "Stephanie wird ein Walross bekochen, und wird permanent den Mann anlächeln." | "late" | "metaphoric" |
| 20 | "Stephanie wird ein Walross bekochen, und wird permanent die Robbe anlächeln." | "late" | "literal" |
| 20 | "Stephanie bekocht ein Walross, und wird permanent den Mann anlächeln." | "early" | "metaphoric" |
| 21 | "Christian umarmt eine Schneeflocke und wird ohne Zweifel die Tochter bewundern." | "early" | "metaphoric" |
| 21 | "Christian umarmt eine Schneeflocke und wird ohne Zweifel die Eiskristalle bewundern." | "early" | "literal" |
| 21 | "Christian wird eine Schneeflocke umarmen und wird ohne Zweifel die Tochter bewundern." | "late" | "metaphoric" |
| 21 | "Christian wird eine Schneeflocke umarmen und wird ohne Zweifel die Eiskristalle bewundern." | "late" | "literal" |
| 22 | "Petra wird einen Zahnstocher ignorieren, und wird unverzüglich den Holzsplitter vergessen." | "late" | "literal" |
| 22 | "Petra ignoriert einen Zahnstocher, und wird unverzüglich den Teenager vergessen." | "early" | "metaphoric" |
| 22 | "Petra ignoriert einen Zahnstocher, und wird unverzüglich den Holzsplitter vergessen." | "early" | "literal" |
| 22 | "Petra wird einen Zahnstocher ignorieren, und wird unverzüglich den Teenager vergessen." | "late" | "metaphoric" |
| 23 | "Lukas wird einen Esel beglückwünschen und wird verständlicherweise den Feuerwehrmann bewundern." | "late" | "metaphoric" |
| 23 | "Lukas wird einen Esel beglückwünschen und wird verständlicherweise das Nutztier bewundern." | "late" | "literal" |
| 23 | "Lukas beglückwünscht einen Esel und wird verständlicherweise den Feuerwehrmann bewundern." | "early" | "metaphoric" |
| 23 | "Lukas beglückwünscht einen Esel und wird verständlicherweise das Nutztier bewundern." | "early" | "literal" |
| 24 | "Jonas favorisiert einen Schrank, und wird schwärmend das Möbelstück betrachten." | "early" | "literal" |
| 24 | "Jonas wird einen Schrank favorisieren, und wird schwärmend den Spieler betrachten." | "late" | "metaphoric" |

| “item number” | “Target sentence” | “Verb constraint” | “context bias” |
|---|---|---|---|
| 24 | “Jonas wird einen Schrank favorisieren, und wird schwärmend das Möbelstück betrachten.” | “late” | “literal” |
| 24 | “Jonas favorisiert einen Schrank, und wird schwärmend den Spieler betrachten.” | “early” | “metaphoric” |
| 25 | “Emily beleidigt ein Fossil und wird von nun an den Professor verpönen.” | “early” | “metaphoric” |
| 25 | “Emily beleidigt ein Fossil und wird von nun an die Versteinerung verpönen.” | “early” | “literal” |
| 25 | “Emily wird ein Fossil beleidigen und wird von nun an den Professor verpönen.” | “late” | “metaphoric” |
| 25 | “Emily wird ein Fossil beleidigen und wird von nun an die Versteinerung verpönen.” | “late” | “literal” |
| 26 | “David wird eine Eisstatue belabern, und wird allmählich die Skulptur in Ruhe lassen.” | “late” | “literal” |
| 26 | “David belabert eine Eisstatue, und wird allmählich die Blondine in Ruhe lassen.” | “early” | “metaphoric” |
| 26 | “David belabert eine Eisstatue, und wird allmählich die Skulptur in Ruhe lassen.” | “early” | “literal” |
| 26 | “David wird eine Eisstatue belabern, und wird allmählich die Blondine in Ruhe lassen.” | “late” | “metaphoric” |
| 27 | “Vanessa wird einen Dschungel begutachten und wird gnadenlos die Wohnung verunglimpfen.” | “late” | “metaphoric” |
| 27 | “Vanessa wird einen Dschungel begutachten und wird gnadenlos den Regenwald verunglimpfen.” | “late” | “literal” |
| 27 | “Vanessa begutachtet einen Dschungel und wird gnadenlos die Wohnung verunglimpfen.” | “early” | “metaphoric” |
| 27 | “Vanessa begutachtet einen Dschungel und wird gnadenlos den Regenwald verunglimpfen.” | “early” | “literal” |
| 28 | “Karoline bewilligt ein Pflaster, und wird gleichzeitig den Wundverband absichern.” | “early” | “literal” |
| 28 | “Karoline wird ein Pflaster bewilligen, und wird gleichzeitig die Arbeitsauszeit absichern.” | “late” | “metaphoric” |
| 28 | “Karoline wird ein Pflaster bewilligen, und wird gleichzeitig den Wundverband absichern.” | “late” | “literal” |
| 28 | “Karoline bewilligt ein Pflaster, und wird gleichzeitig die Arbeitsauszeit absichern.” | “early” | “metaphoric” |

| "item number" | "Target sentence" | "Verb constraint" | "context bias" |
|---|---|---|---|
| 29 | "Laura coacht einen Leoparden und wird mit Vergnügen den Schüler unterrichten." | "early" | "metaphoric" |
| 29 | "Laura coacht einen Leoparden und wird mit Vergnügen die Raubkatze unterrichten." | "early" | "literal" |
| 29 | "Laura wird einen Leoparden coachen und wird mit Vergnügen den Schüler unterrichten." | "late" | "metaphoric" |
| 29 | "Laura wird einen Leoparden coachen und wird mit Vergnügen die Raubkatze unterrichten." | "late" | "literal" |
| 30 | "Philipp wird eine Schlange lobpreisen, und wird unüberlegt dem Kriechtier vertrauen." | "late" | "literal" |
| 30 | "Philipp lobpreist eine Schlange, und wird unüberlegt der Hausfrau vertrauen." | "early" | "metaphoric" |
| 30 | "Philipp lobpreist eine Schlange, und wird unüberlegt dem Kriechtier vertrauen." | "early" | "literal" |
| 30 | "Philipp wird eine Schlange lobpreisen, und wird unüberlegt der Hausfrau vertrauen." | "late" | "metaphoric" |
| 31 | "Nadine wird einen Panzer verärgern, und wird unverzüglich dem Vorgesetzten entfliehen." | "late" | "metaphoric" |
| 31 | "Nadine wird einen Panzer verärgern, und wird unverzüglich dem Kampfwagen entfliehen." | "late" | "literal" |
| 31 | "Nadine verärgert einen Panzer, und wird unverzüglich dem Vorgesetzten entfliehen." | "early" | "metaphoric" |
| 31 | "Nadine verärgert einen Panzer, und wird unverzüglich dem Kampfwagen entfliehen." | "early" | "literal" |
| 32 | "Sebastian füttert eine Prinzessin, und wird unablässig der Adligen beistehen." | "early" | "literal" |
| 32 | "Sebastian wird eine Prinzessin füttern, und wird unablässig der Katze beistehen." | "late" | "metaphoric" |
| 32 | "Sebastian wird eine Prinzessin füttern, und wird unablässig der Adligen beistehen." | "late" | "literal" |
| 32 | "Sebastian füttert eine Prinzessin, und wird unablässig der Katze beistehen." | "early" | "metaphoric" |
| 33 | "Fabian erzieht einen Papagei, und wird selbstverständlich dem Jungen zujubeln." | "early" | "metaphoric" |
| 33 | "Fabian erzieht einen Papagei, und wird selbstverständlich dem Vogel zujubeln." | "early" | "literal" |
| 33 | "Fabian wird einen Papagei erziehen, und wird selbstverständlich dem Jungen zujubeln." | "late" | "metaphoric" |

| **“item number”** | **“Target sentence”** | **“Verb constraint”** | **“context bias”** |
|---|---|---|---|
| 33 | “Fabian wird einen Papagei erziehen, und wird selbstverständlich dem Vogel zujubeln.” | “late” | “literal” |
| 34 | “Alexander wird einen Piranha belehren, und wird furchtlos dem Fisch entgegentreten.” | “late” | “literal” |
| 34 | “Alexander belehrt einen Piranha, und wird furchtlos dem Teilnehmer entgegentreten.” | “early” | “metaphoric” |
| 34 | “Alexander belehrt einen Piranha, und wird furchtlos dem Fisch entgegentreten.” | “early” | “literal” |
| 34 | “Alexander wird einen Piranha belehren, und wird furchtlos dem Teilnehmer entgegentreten.” | “late” | “metaphoric” |
| 35 | “Simone wird einen Geier protegieren, und wird sich wiederholt für den Ziehsohn einsetzen.” | “late” | “metaphoric” |
| 35 | “Simone wird einen Geier protegieren, und wird sich wiederholt für den Greifvogel einsetzen.” | “late” | “literal” |
| 35 | “Simone protegiert einen Geier, und wird sich wiederholt für den Ziehsohn einsetzen.” | “early” | “metaphoric” |
| 35 | “Simone protegiert einen Geier, und wird sich wiederholt für den Greifvogel einsetzen.” | “early” | “literal” |
| 36 | “Niklas umwirbt eine Trophäe, und wird unbeirrt den Pokal anvisieren.” | “early” | “literal” |
| 36 | “Niklas wird eine Trophäe umwerben, und wird unbeirrt die Sängerin anvisieren.” | “late” | “metaphoric” |
| 36 | “Niklas wird eine Trophäe umwerben, und wird unbeirrt den Pokal anvisieren.” | “late” | “literal” |
| 36 | “Niklas umwirbt eine Trophäe, und wird unbeirrt die Sängerin anvisieren.” | “early” | “metaphoric” |

C

# Appendix C

## C.1 Pilot Study of Experiment 5

Before conducting Experiment 5, a pilot study was run with 12 participants in order to (1) be able to make more precise predictions regarding the possible outcome of the main experiment and (2) determine the number of participants necessary for the study to have more than 80% power.

The original analysis scheme was set up as can be seen in Table C.1 below. The time-windows (regions) for analysis were determined by focusing on a comparison of what the first and second metaphoric element would be in each condition. This resulted in Region 1 comparing the verb (early-verb condition) and vehicle (late verb condition) and Region 2 the vehicle (early-verb condition and the verb (late verb condition). I report the results of the pilot study using this original analysis, as well as the results of regions 1 and 2 of Experiment 5. Regions 3-5 are reported in Chapter 5, since no changes were made here.

**Table C.1:** Original region distribution of Experiment 5

| Region | Definition | Example region, early Verb condition | Example region, late verb Condition |
|---|---|---|---|
| 1 | from onset1 to onset2 | füttert | eine Prinzessin |

| Region | Definition | Example region, early Verb condition | Example region, late verb Condition |
|---|---|---|---|
| 2 | from onset 2 to onset "und" | eine Prinzessin | füttern |
| 3 | from onset "und" to onset "adv" | und wird | und wird |
| 4 | from from onset "adv" to onset "dis" | unablässig | unablässig |
| 5 | from onset dis to end of sentence | der Adligen *literal*, der Katze *metaphoric* beistehen | der Adligen *literal*, der Katze *metaphoric* beistehen |

Another crucial difference was the dependent measure used for analysis. Instead of examining log-gaze ratios of looks to the metaphoric picture divided by looks to the literal picture, I originally took the log-gaze ratio of looks to the target picture divided by looks to the competitor picture. Target and competitor varied between conditions: The metaphoric picture was the target and the literal picture was the competitor in the metaphoric conditions and the opposite was the case in the literal conditions.

### C.1.1 Participants

The pilot study was conducted with 12 participants (7 female) ages 18-31. One of the participants was removed from the analysis prior to data inspection based on the results of their post-experiment questionnaire. This participant was replaced, keeping the totla number at 12. None of the participants of this pilot study participated in any of the norming tasks or in Experiments 5 or 6. All participants were right handed and had normal or corrected-to-normal vision. This study was approved by the ethics vote of the psycholinguistics lab of the Humboldt-Universität zu Berlin.

### C.1.2 Predictions

The first set of predictions referred to the effect of context on the gaze record. If the contextually-derived mental representations affect participants' eye-gaze differently in the two different context bias conditions (literal and metaphoric), we should find that there is an overall preference for inspecting the target image across conditions (i.e. the metaphoric picture in the metaphoric conditions and the literal picture in the literal conditions).

This could, however, be modulated by the way that context interacts with the selection restrictions of the verb in the critical sentence: In the early verb conditions, it could be the case that the verb's selectional preferences are stronger than contextual biases. If this is the case, then we should find a difference in log-gaze ratios between literal and metaphoric context bias levels of the early verb condition: This would suggest that in the literal condition, participants looked more at the picture most associated with the verb (the metaphoric picture, competitor image) than at the one most associated with the context (the literal picture, target image).

The second set of predictions referred to the effect of the position of the verb in the sentence. Broadly speaking, if presentation order of the metaphoric elements is important for the construction of metaphoric meaning (as claimed by class inclusion models), we would expect participants to settle on a figurative interpretation sooner when they first hear the verb and then the direct object than when presentation order is reversed. This could translate to differences in the time in which participants establish reference with the target image (metaphoric picture) in the metaphoric conditions: It could be the case that the point in time in which log-gaze ratios become positive will occur during region 2 for the metaphoric early verb condition, whereas this could happen in later regions only for the metaphoric late-verb condition.

### C.1.3 Results

The results are shown in Figures C.1, C.2, C.3, C.4 and C.5. The results of the models, shown in Table C.2 show an overall preference for the target picture in all conditions with the exception of the late-verb metaphoric condition in region 1, which shows a slight preference for the competitor, literal picture. This pattern suggests that participants settled on a metaphoric interpretation of the sentence more easily in the early-verb compared to late-verb metaphoric conditions: In region 1, upon hearing the metaphoric vehicle **Prinzessin** (late-verb metaphoric condition), participants showed a preference for the competitor literal picture (the princess) that resulted in a large difference in log-ratios compared to the literal late-verb condition. In region 2, on the other hand, upon hearing the metaphoric vehicle (early-verb metaphoric condition), participants showed a slight preference for the metaphoric picture (the cat), which resulted in a less pronounced difference in log-ratios compared to the literal early condition.

Concretely, this differential pattern in the two regions translated into an interaction between the two factors in region 1: The difference between log-ratios in the late-verb conditions was significantly larger than the difference between log-ratios in the early conditions. Conversely, there was no significant interaction inn region 2: The difference between log-ratios in the late-verb metaphoric condition was not significantly larger than the difference between the early-verb conditions.

On the other hand, the log-ratio for both early-verb conditions was positive in region 1, suggesting that contextual expectations might be stronger than verb expectations: Participants preferentially inspected the target picture in both conditions, which was compatible with the verb's selectional preferences in the metaphoric conditions, but incompatible with these in the literal conditions. However, this preference was more pronounced in the metaphoric condition, which might indicate that verb-driven expectations do interfere to a certain degree with contextually-driven expectations: Contextual and verb expectation are aligned in the early-verb metaphoric condition, encouraging participants to look towards the target, whereas contextual and verb-driven expectations were at odds in the early-verb literal condition, potentially causing a momentary uncertainty, which translated into lower proportion of looks to target.

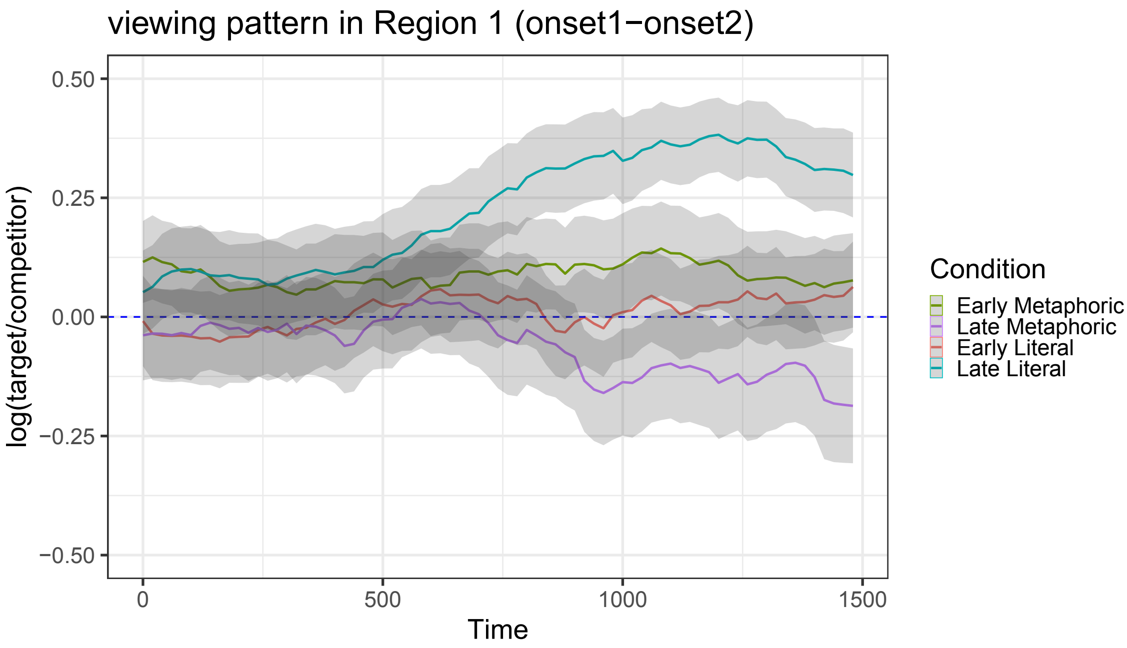


**Figure C.1:** Original Results of Region 1, pilot study of Experiment 5

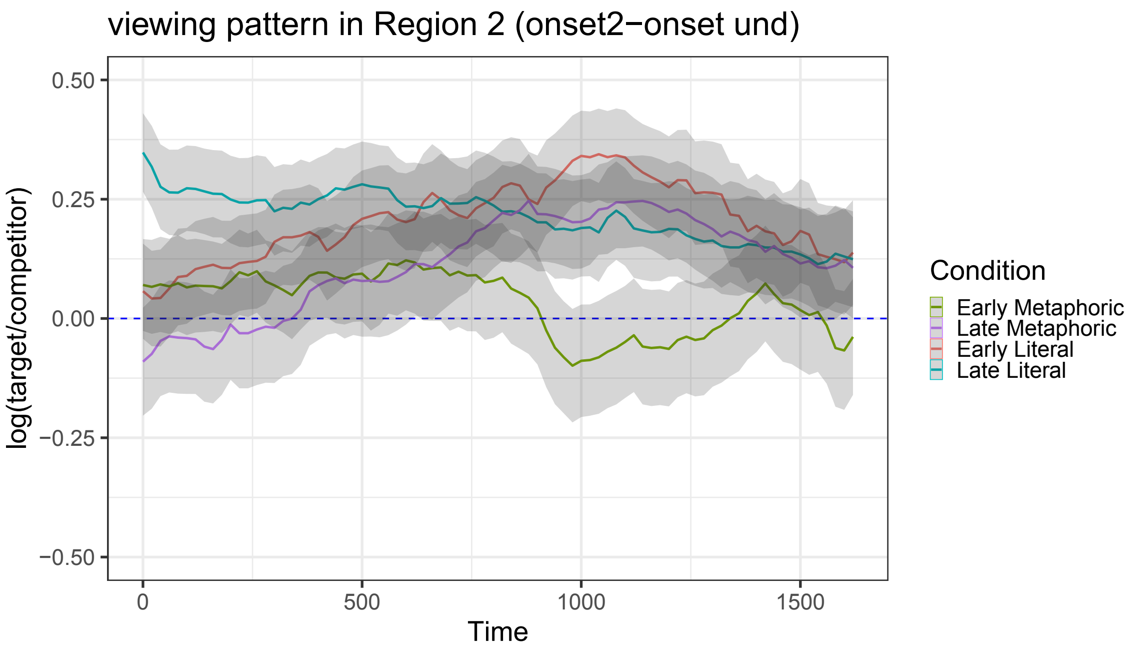


**Figure C.2:** Original Results of Region 2, pilot study of Experiment 5

The results for the remaining 3 regions showed that participants quickly identified the target image and settled on this until the end of the trial: Overall, the intercept of every model was positive and significantly different from zero, meaning that there as an overall preference for target picture throughout these three regions.

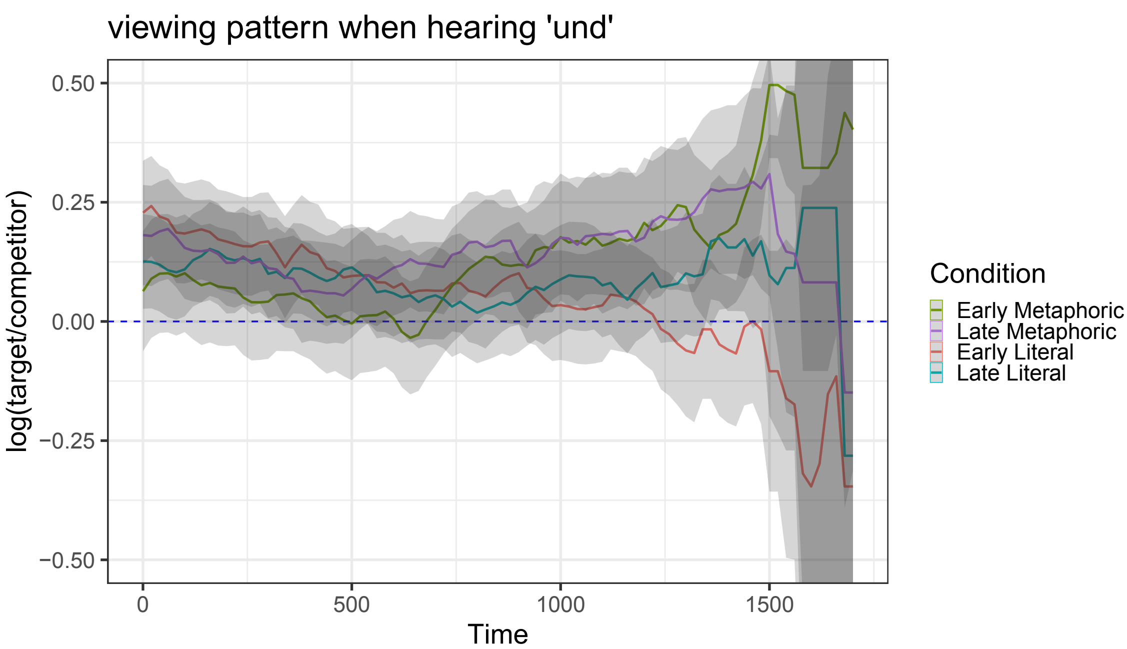


**Figure C.3:** Original Results of Region 3, pilot study of Experiment 5

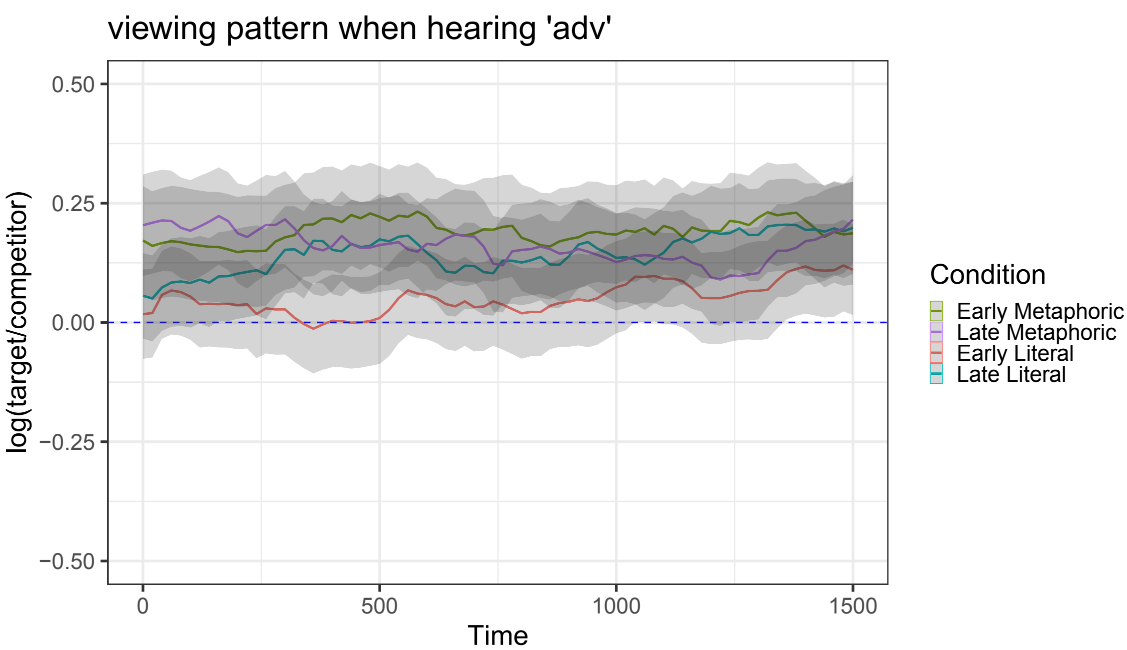


**Figure C.4:** Original Results of Region 4, pilot study of Experiment 5

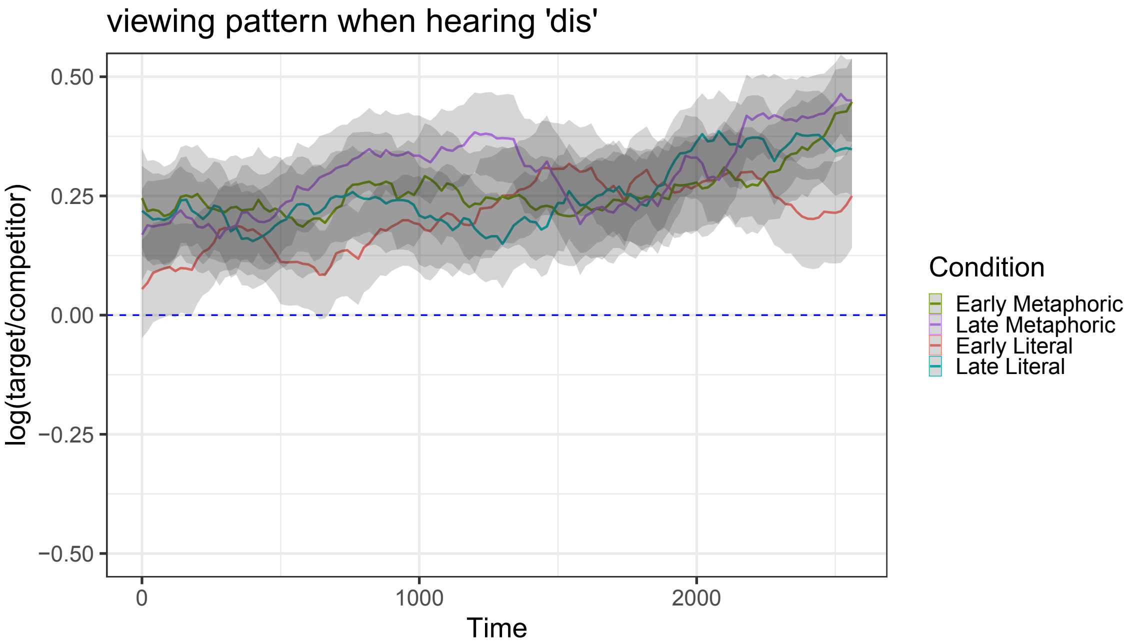


**Figure C.5:** Original Results of Region 5, pilot study of Experiment 5

**Table C.2:** Regression results for regions 1-5, pilot study of Experiment 5

| | *Dependent variable:* | | | | |
|---|---|---|---|---|---|
| | logratio of looks to target/looks to competitor | | | | |
| | (1) | (2) | (3) | (4) | (5) |
| Context Bias | −0.069<br>(−0.215, 0.077)<br>t = −0.922 | −0.134<br>(−0.249, −0.020)<br>t = −2.298* | 0.056<br>(−0.093, 0.205)<br>t = 0.734 | 0.106<br>(−0.033, 0.245)<br>t = 1.499 | 0.082<br>(−0.069, 0.234)<br>t = 1.064 |
| Verb constraint | 0.042<br>(−0.068, 0.153)<br>t = 0.749 | −0.059<br>(−0.162, 0.043)<br>t = −1.131 | −0.030<br>(−0.180, 0.120)<br>t = −0.390 | −0.040<br>(−0.162, 0.082)<br>t = −0.647 | −0.053<br>(−0.156, 0.050)<br>t = −1.013 |
| Trial Order | 0.001<br>(−0.0001, 0.001)<br>t = 1.665 | 0.003<br>(0.002, 0.003)<br>t = 8.667*** | −0.00003<br>(−0.001, 0.001)<br>t = −0.069 | 0.005<br>(0.004, 0.005)<br>t = 15.190*** | 0.002<br>(0.001, 0.002)<br>t = 6.794*** |
| Interaction | −0.371<br>(−0.509, −0.233)<br>t = −5.271*** | −0.035<br>(−0.310, 0.240)<br>t = −0.247 | −0.100<br>(−0.283, 0.082)<br>t = −1.079 | 0.145<br>(−0.027, 0.317)<br>t = 1.649 | −0.007<br>(−0.024, 0.011)<br>t = −0.730 |
| Intercept | 0.054<br>(0.041, 0.067)<br>t = 8.086*** | 0.080<br>(0.066, 0.093)<br>t = 11.865*** | 0.101<br>(0.086, 0.116)<br>t = 13.455*** | 0.045<br>(0.032, 0.058)<br>t = 6.917*** | 0.224<br>(0.214, 0.234)<br>t = 44.634*** |
| Analysis region | 1 | 2 | 3 | 4 | 5 |
| Observations | 29,700 | 32,472 | 24,875 | 30,096 | 51,084 |
| Log Likelihood | −20,589.470 | −24,094.700 | −18,230.870 | −20,296.230 | −35,610.500 |
| Akaike Inf. Crit. | 41,200.930 | 48,211.400 | 36,483.750 | 40,614.450 | 71,241.010 |
| Bayesian Inf. Crit. | 41,292.220 | 48,303.670 | 36,573.080 | 40,705.890 | 71,329.420 |
| *Note:* | *p<0.05; **p<0.01; ***p<0.001<br>values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) | | | | |

## C.2 Changes Made Prior to Conducting Experiment 5

Based on the results and analysis of the pilot study, some changes were made prior to conducting the main Experiment.

The first change concerned the type of fillers used in the experiment. By comparing the gaze patterns of participants during the first and second half of the pilot study a pattern appeared suggesting that participants developed an adaption strategy throughout the study. During the first half of the experiment, the literal picture attracted more looks in the metaphoric conditions upon hearing the metaphoric vehicle (in region 1 for the late-verb and in region 2 for the early-verb condition) than in the second half of the experiment. This could be due to participants anticipating the correct target image based solely on the information provided by the linguistic context: In all critical and almost all filler trials, the information about which picture to select did not come from the spoken sentence but from the written text. Participants could have therefore learned to suppress the incoming information from the sentence and focus only on what they had read.

To address this issue, 36 of the filler items were changed in order to make sure that the information needed to decide which picture to click on came exclusively from the spoken sentence and not from the written context. This should force participants to pay close attention to the sentences and not settle on an interpretation based on the context alone.

Further changes concerned the analysis scheme. In order to simplify the analysis and interpretation, I decided to change the dependent measure from log(target/competitor) to log(metaphoric picture/literal picture). This should facilitate the interpretation of the comparison of results in the metaphoric and literal conditions.

Furthermore, I decided to pot for a different contrast coding scheme to analyze the data from Experiment 5: Instead of using an ANOVA-style sum-contrast coding scheme, I chose to instead use treatment contrast coding. The logic for this was as follows: The results of the pilot study suggested a different pattern of construction of metaphoric meaning in the early vs late metaphoric conditions: Whereas in the late-verb condition, upon hearing the metaphoric vehicle **Prinzessin** (region 1) participants showed an overall preference for looking at the literal picture (the princess), such preference was not visible when the same word was heard in the early-verb conditions (i.e. in region 2). Here, participants showed a preference for the target metaphoric picture (the cat).

One way to quantify this difference is to tailor the statistical model to the following question:

*(i) When participants hear the metaphorical vehicle, do they preferentially look at the literal picture, at the metaphoric picture, or do they show no preference at all?*

This question can be directly addressed by taking advantage of the flexibility in contrast coding offered by linear models, which is fundamentally different to the one used in an analysis of variance analysis (ANOVA) (Schad, 2020; Gelman & Hill, 2007; Barr et al., 2013). When using linear models with categorical predictors, the experimenter must specify a baseline to which all of the experimental groups are compared, which is known as the process of **contrast coding**. One such contrast coding strategy is the so-called treatment or "dummy" contrast, in which one specific group is used as a baseline, with the coefficients of the predictor variables representing direct comparisons between a predictor and the baseline group while keeping all other predictors at a fixed level. This differs from the sum-contrast used for the pilot study (which most resembles the contrast coding scheme used in ANOVA-style analyses), in which the baseline represents the overall mean of all conditions, and the coefficients of the predictor variables represent the effect of a specific factor compared to the overall mean.

Importantly, when using treatment contrast the coefficient of the intercept tests the null-hypothesis of whether the outcome value (i.e. log-gaze probability ratio) for the group chosen as the intercept differs significantly from zero or not. Recall, that log-gaze probability ratios are centered around zero, with positive numbers signifying a preference for the metaphoric picture and negative numbers a preference for the literal picture. This means that interpreting the intercept term in a linear model using treatment contrast will tell us if for the specific group we have chosen as the baseline (i.e the intercept) of the model there is a looking preference for one or the other picture.

In region 1, participants heard the vehicle of a metaphor in the late-verb metaphoric condition, whereas in region 2 this happens in the early-verb metaphoric condition. Thus, if we code the metaphoric early-verb and the metaphoric late-verb conditions as the baseline conditions for the statistical models of region 1 and region 2 respectively, we will be able to directly answer question (i): If the intercept coefficient is significantly positive, it will mean that participants had an overall preference for looking at the metaphoric picture (the cat) upon hearing the metaphoric vehicle (**Prinzessin**); if the intercept coefficient is significantly negative, it will mean that participants preferred to look at the literal picture (the princess) upon hearing the metaphoric vehicle.

The interpretation of the coefficients using treatment contrast is different than the interpretation when using sum contrast for all effects except for the highest order one (i.e. the interaction term) (Levy, 2014). This means that care must be taken when formulating the predictions and analyzing the results. What follows is a reformulation of the predictions for the main experiment considering the findings of the pilot study and the new coding scheme for the statistical analysis.

## C.3 Predictions for Experiment 5 with the Original Analysis Scheme

### C.3.1 Region 1

The main prediction here was that when coding the late-verb metaphoric condition as the intercept of the model, this should be negative and potentially also significantly different from zero. This would reflect that participants either preferentially inspected the literal picture (if significantly different from zero) or that they were uncertain about which image was the correct referent (if not different from zero).

Furthermore, there should be significant differences between metaphoric and literal conditions: Both metaphoric conditions should have significantly more positive log-gaze ratios compared to their literal counterparts, signifying that participants spent more time looking at the metaphoric picture than at the literal one in the metaphoric conditions.

### C.3.2 Region 2

I predicted that, when coding the early-verb metaphoric condition as the intercept, this should be positive and significantly different than zero. This would reflect an overall preference for the metaphoric picture in this condition. There should also be differences between corresponding literal and metaphoric conditions.

## C.4 Determining the Number of Participants Needed for Experiment 5

In order to determine the number of participants for the main experiment, a simulation approach was used, following the recommendations of Vasishth et al. (2018) as well as Green and MacLeod (2016). This was done using the R package Simr (Green & MacLeod, 2016). Using the results of the maximal linear mixed-effects model for region 1, 100 simulated experiments were run with with artificial data. This data was created using the extracted parameters obtained from the statistical model fitted to region 1. Crucially, these parameters were slightly altered prior to the creation of the artificial data sets: The interaction effect, which had a value of Cohens'd of 0.38, was reduced to one with the value of Cohen's d of 0.2. This way, I intended to find the number of participants needed to have at least 80% power assuming a more conservative effect size. The simulations suggest that with 32 participants, the experiment should be sufficiently powered to achieve this goal.

Given that the analysis scheme was also changed after the results of Experiment 5 were known (see the final section of this appendix), this power analysis was re-run to ensure that Experiment 5 was sufficiently powered to detect a difference between early and late metaphoric conditions when the data was realigned to the onset of the metaphoric vehicle (as described in chapter four). Again, the effect size for this specific difference in the pilot study was changed to one of 0.2 (in Cohen's d units) prior to conducting the simulations. The simulations showed that, based on the results of the pilot study, Experiment 5 would also be sufficiently powered to find an effect of at least 0.2 (in Cohen's d units) with 32 participants.

## C.5 Results of Experiment 5 with the Original Analysis

### C.5.1 Region 1

Three versions of the same model were fitted to the data, coding the early metaphoric, late metaphoric and early literal conditions as the intercept respectively. The models show that the early metaphoric condition was positive and significantly different from zero. This condition was also significantly different from the early-literal condition and from the metaphoric-late condition. These results can be visualized in Figure C.6 and in Table C.3.

When coding the late-verb metaphoric as the intercept, this was found to be negative and significantly different from zero. This means that participants looked more at the literal picture than at the metaphoric in this condition throughout this region.

Participants also looked significantly more to the literal picture in the literal late condition than in the metaphoric late condition, given the negative and significant difference between these conditions.

### C.5.2 Region 2

Three version of the same model were fitted to the data, coding the early metaphoric, late metaphoric and early literal conditions as the intercept respectively. The models show that the early metaphoric condition was positive and significantly different from zero. This condition was also significantly different from the early-literal condition and from the metaphoric-late condition. When coding the late-verb

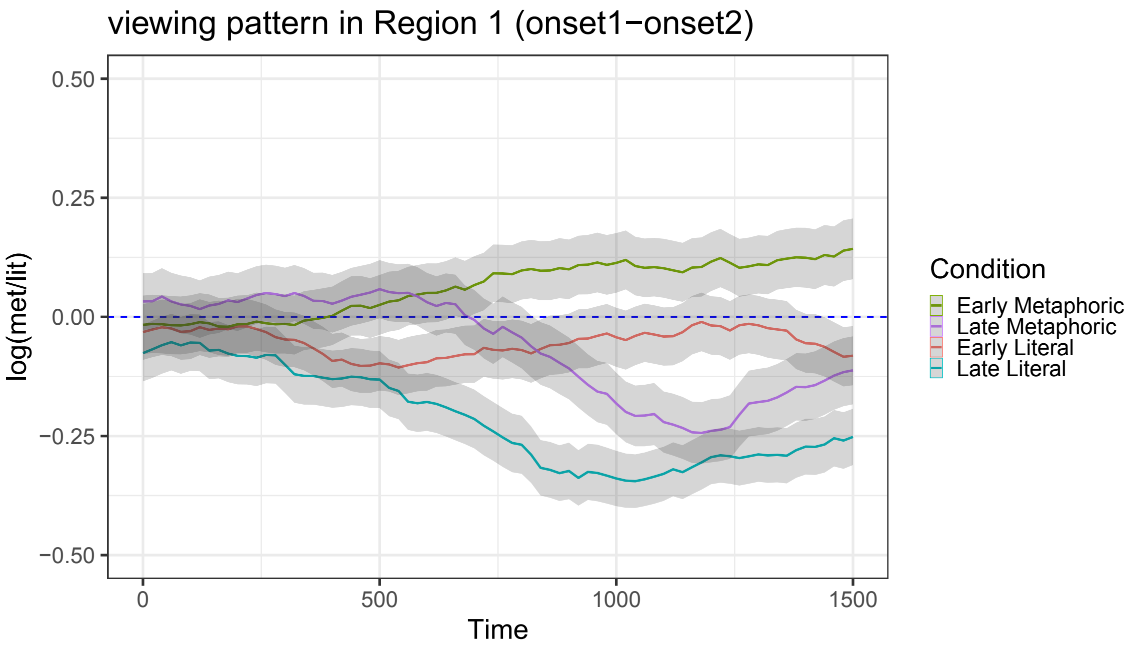


**Figure C.6:** Original Results of Region 1, Experiment 5

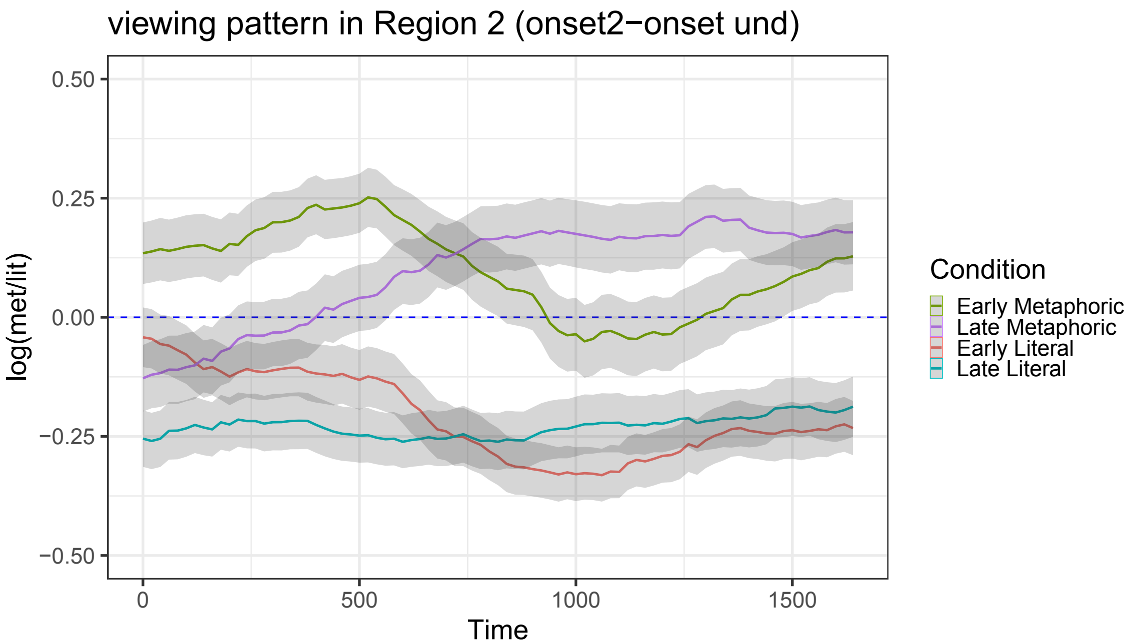


**Figure C.7:** Original Results of Region 2, Experiment 5

metaphoric as the intercept, this was found to also be positive but not significantly different from zero. These results can be visualized in Figure C.7 and in Table C.4.

**Table C.3:** Original regression analysis of log-gaze ratio in Region 1 of Experiment 5

| | *Dependent variable:* | | |
|---|---|---|---|
| | logratio | | |
| | (1) | (2) | (3) |
| Early-Metaphoric vs. Early-Literal | −0.116<br>(−0.218, −0.014)<br>t = −2.236* | | |
| Early-metaphoric vs. Late-metaphoric | −0.127<br>(−0.207, −0.047)<br>t = −3.124** | | |
| Late-metaphoric vs. late-literal | | −0.141<br>(−0.151, −0.131)<br>t = −28.128*** | |
| Early-literal vs. Late-literal | | | −0.163<br>(−0.225, −0.101)<br>t = −5.159*** |
| Trial Order | −0.001<br>(−0.001, −0.0004)<br>t = −3.902*** | −0.0004<br>(−0.001, −0.00005)<br>t = −2.235* | −0.0005<br>(−0.001, −0.0001)<br>t = −2.651** |
| verb/context interaction | −0.034<br>(−0.141, 0.072)<br>t = −0.636 | | |
| Intercept | 0.075<br>(0.007, 0.144)<br>t = 2.166* | −0.063<br>(−0.111, −0.015)<br>t = −2.580** | −0.045<br>(−0.084, −0.006)<br>t = −2.261* |
| condition at intercept | early-met | late-met | early-lit |
| Observations | 82,536 | 82,536 | 82,536 |
| Log Likelihood | −57,408.830 | −60,661.300 | −61,459.250 |
| Akaike Inf. Crit. | 114,845.700 | 121,342.600 | 122,936.500 |
| Bayesian Inf. Crit. | 114,976.200 | 121,435.800 | 123,020.400 |

*Note:* *p<0.05; **p<0.01; ***p<0.001
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table C.4:** Original regression analysis of log-gaze ratio in Region 2 of Experiment 5

| | *Dependent variable:* | | |
|---|---|---|---|
| | logratio | | |
| | (1) | (2) | (3) |
| Early-Metaphoric vs. Early-Literal | −0.313<br>(−0.394, −0.231)<br>t = −7.512*** | | |
| Early-metaphoric vs. Late-metaphoric | −0.013<br>(−0.077, 0.051)<br>t = −0.394 | | |
| Late-metaphoric vs. late-literal | | −0.309<br>(−0.390, −0.227)<br>t = −7.419*** | |
| Early-literal vs. Late-literal | | | −0.009<br>(−0.073, 0.055)<br>t = −0.278 |
| Trial Order | 0.002<br>(0.002, 0.002)<br>t = 10.549*** | 0.002<br>(0.002, 0.002)<br>t = 10.549*** | 0.002<br>(0.002, 0.002)<br>t = 10.549*** |
| verb/context interaction | 0.004<br>(−0.010, 0.018)<br>t = 0.534 | | |
| Intercept | 0.065<br>(0.008, 0.123)<br>t = 2.232* | 0.053<br>(−0.005, 0.110)<br>t = 1.792 | −0.247<br>(−0.304, −0.190)<br>t = −8.440*** |
| condition at intercept | early-met | late-met | early-lit |
| Observations | 90,138 | 90,138 | 90,138 |
| Log Likelihood | −70,868.620 | −70,868.620 | −70,868.620 |
| Akaike Inf. Crit. | 141,759.200 | 141,759.200 | 141,759.200 |
| Bayesian Inf. Crit. | 141,862.700 | 141,862.700 | 141,862.700 |

*Note:* *p<0.05; **p<0.01; ***p<0.001
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

### C.5.3 Discussion

When comparing regions 1 and 2 it is possible to obtain indirect evidence supporting an asymmetric processing of metaphoric meaning: When hearing the metaphoric vehicle (**Prinzessin**), participants' viewing pattern was different as a function of whether or not they had previously heard the verbal topic: When they had heard it (region 2, early-verb metaphoric condition), participants preferentially inspected the metaphoric picture throughout the region. When they hadn't heard it yet (region 1, late-verb metaphoric condition), participants preferentially inspected the literal picture throughout the region.
Furthermore, both regions showed differences between metaphoric and literal conditions: as a function of contextual bias, participants preferred to inspect the metaphoric picture in the metaphoric conditions and the literal picture in the literal conditions.

### C.5.4 Reasoning behind changing the analysis

This interpretation of the results is tentative and must be taken with care: For one, it is based on a comparison between the analyses of regions 1 and 2, which are, in turn, the result of comparing gaze patterns when hearing two different words: the verb and the object of a sentence. This is less than ideal because it leaves us only with indirect evidence for our main hypothesis, which is that a metaphor will be processed differently as a function of the position of the elements. In order to test this hypothesis in a more direct way, it would be better to compare participants' looking behavior across conditions upon hearing *the same word*. This could be achieved by re-aligning the time-course of participants' viewing patterns to the onset of the metaphoric vehicle (the word **Prinzessin**) and to the onset of the verb (the word **füttert**).

This way of analyzing the data has the big advantage that it allows for a direct comparisons of differences in processing the same words in order to analyze how their processing differs as a function of the word's position in the sentence. This is the final analysis that is reported in chapter four.

# D

# Appendix D

## D.1 Complementary Analyses for Experiment 6

Experiment 6 focused on the results for the Vehicle and Verb regions. The remaining regions (UND, ADV, and DIS) were analyzed analogously to the Verb and Vehicle regions: Models including verb constraint as a fixed effect as well as trial order as a control were fitted to the log-gaze probability ratios of looks to metaphoric picture over looks to literal picture. All models used a treatment contrast coding scheme. They were each fitted twice, changing the intercept of the model from the early to the late verb constraint condition respectively.

For the DIS region, the model included the factor "Disambiguation", with the levels "metaphoric" and "literal", in addition to the factor "verb constraint". This is because for this region only, there were differences in the information given to participants about the likely visual referent: In the literal conditions, participants heard the literal word ("noble woman") and in the metaphoric conditions participants heard the metaphoric word ("cat").

### D.1.1 Results of Experiment 6

#### UND region

Results from the “und” region showed an overall preference for the literal over the metaphoric picture. This is shown by the fact that both conditions displayed a negative log-gaze ratio that was significantly different from zero, meaning that, in both conditions, participants preferred to look at the literal picture than at the metaphoric picture. There was no significant difference in viewing behavior between conditions in this region. These results are displayed in Figure D.1 and the model output is shown in Table D.1.

#### ADV region

Similarly to the results of the previous region, both conditions were negative and significantly different from zero, indicating a preference for the literal picture throughout the region. There was no difference between the two conditions. These results are displayed in Figure D.2 and the model output is shown in Table D.2.

#### DIS region

The analysis for this region was different than the previous analyses since it included the additional two-level factor Disambiguation (literal vs. metaphoric). The model for this region thus included the two factors (verb constraint and disambiguation) as predictors, together with their interaction. The data was therefore analyzed in the same way as it was in Experiment 5.

The results were somewhat similar to those found in Experiment 5 for this same time region, as can be seen in Figure D.3: Participants directed their gaze towards the picture representing the word heard in the disambiguating region. This translated to significant effects of disambiguation for both early and late verb constraint conditions. Furthermore, there was also a significant difference between early and late metaphoric conditions: Participants looked more at the literal picture compared to the metaphoric picture in the metaphoric late verb condition compared to the metaphoric early-verb condition. However, as can be seen by the raw effect sizes

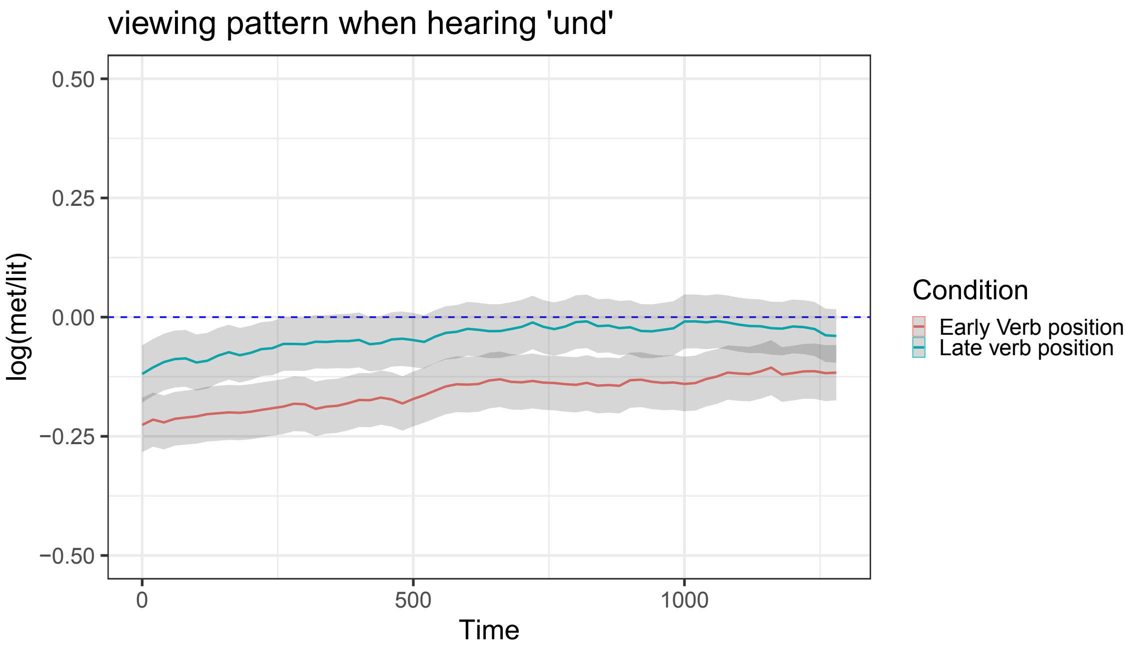


**Figure D.1:** Summary of results for the UND region, Experiment 6

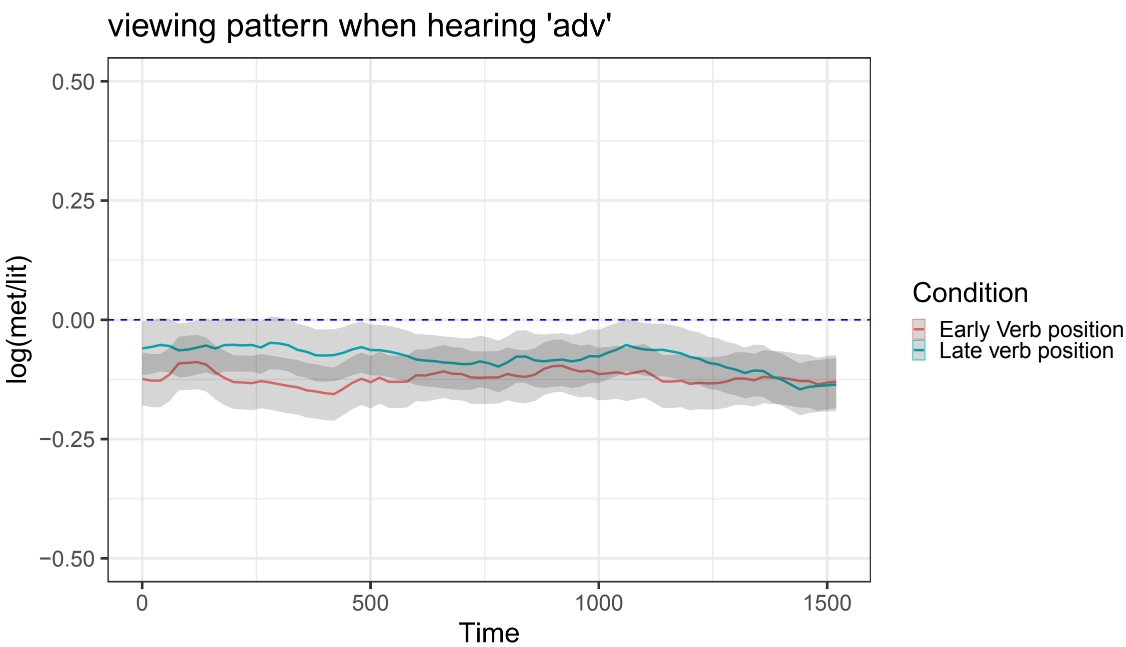


**Figure D.2:** Summary of results for the ADV region, Experiment 6

reported in Table D.3, this effect was much smaller than the effect of Disambiguation. The intercept in the metaphoric conditions was positive and significantly different from zero, while the intercepts in the literal conditions was negative and significantly different from zero. All model outputs are shown in Table D.3.

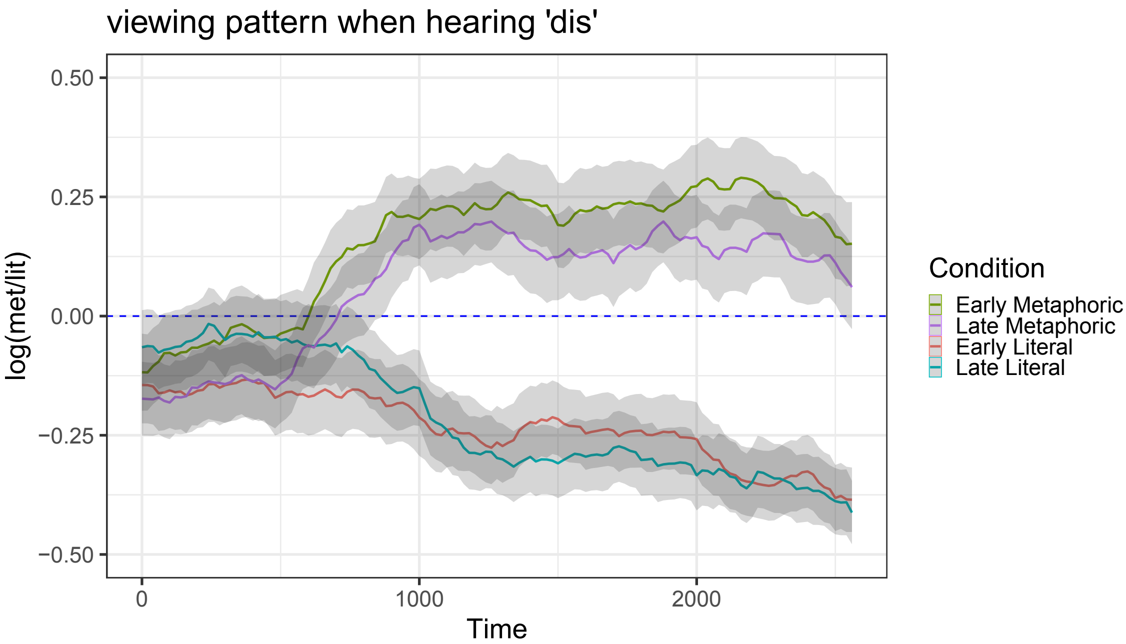


**Figure D.3:** Summary of results for the DIS region, Experiment 6

**Table D.1:** Regression analysis of log-gaze probability ratios in the 'und' region of Experiment 6

| | *Dependent variable:* | |
|---|---|---|
| | logratio | |
| | (1) | (2) |
| Early-Metaphoric vs. Early-Literal | 0.074 | |
| | (−0.039, 0.186) | |
| | t = 1.281 | |
| Trial Order | 0.003 | 0.003 |
| | (0.002, 0.003) | (0.002, 0.003) |
| | t = 11.405*** | t = 11.405*** |
| Intercept | −0.191 | −0.117 |
| | (−0.248, −0.133) | (−0.174, −0.060) |
| | t = −6.533*** | t = −3.990*** |
| condition at intercept | early-met | late-met |
| Observations | 46,930 | 46,930 |
| Log Likelihood | −39,411.490 | −39,411.490 |
| Akaike Inf. Crit. | 78,834.970 | 78,834.970 |
| Bayesian Inf. Crit. | 78,887.510 | 78,887.510 |

*Note:* *p<0.05; **p<0.01; ***p<0.001
values shown per cell are regression coefficients, confidence intervals and t-values ( in that order)

**Table D.2:** Regression analysis of log-gaze probability ratios in the 'adv' region of Experiment 6

| | *Dependent variable:* | |
|---|---|---|
| | logratio | |
| | (1) | (2) |
| Early-Metaphoric vs. Early-Literal | 0.003<br>(−0.096, 0.102)<br>t = 0.059 | |
| Trial Order | 0.003<br>(0.003, 0.003)<br>t = 12.901*** | 0.003<br>(0.003, 0.003)<br>t = 12.901*** |
| Intercept | −0.158<br>(−0.223, −0.093)<br>t = −4.762*** | −0.155<br>(−0.215, −0.095)<br>t = −5.062*** |
| condition at intercept | early-met | late-met |
| Observations | 55,594 | 55,594 |
| Log Likelihood | −43,987.670 | −43,987.670 |
| Akaike Inf. Crit. | 87,991.340 | 87,991.340 |
| Bayesian Inf. Crit. | 88,062.750 | 88,062.750 |
| *Note:* | *p<0.05; **p<0.01; ***p<0.001<br>values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) | |

**Table D.3:** Regression analysis of log-gaze probability ratios in the 'dis' region of Experiment 6

| | *Dependent variable:* | | |
|---|---|---|---|
| | logratio | | |
| | (1) | (2) | (3) |
| Early-Metaphoric vs. Early-Literal | −0.387<br>(−0.442, −0.333)<br>t = −14.025*** | | |
| Early-metaphoric vs. Late-metaphoric | −0.090<br>(−0.100, −0.080)<br>t = −17.256*** | | |
| Late-metaphoric vs. late-literal | | −0.282<br>(−0.336, −0.228)<br>t = −10.197*** | |
| Early-literal vs. Late-literal | | | 0.016<br>(0.006, 0.026)<br>t = 3.085** |
| Trial Order | 0.001<br>(0.001, 0.002)<br>t = 6.348*** | 0.001<br>(0.001, 0.002)<br>t = 6.348*** | 0.001<br>(0.001, 0.002)<br>t = 6.348*** |
| verb*context interaction | 0.106<br>(0.091, 0.120)<br>t = 14.426*** | | |
| Intercept | 0.142<br>(0.078, 0.206)<br>t = 4.357*** | 0.052<br>(−0.012, 0.116)<br>t = 1.598 | −0.246<br>(−0.309, −0.182)<br>t = −7.541*** |
| condition at intercept | early-met | late-met | early-lit |
| Observations | 93,136 | 93,136 | 93,136 |
| Log Likelihood | −77,977.010 | −77,977.010 | −77,977.010 |
| Akaike Inf. Crit. | 155,970.000 | 155,970.000 | 155,970.000 |
| Bayesian Inf. Crit. | 156,045.500 | 156,045.500 | 156,045.500 |
| *Note:* | *p<0.05; **p<0.01; ***p<0.001<br>values shown per cell are regression coefficients, confidence intervals and t-values ( in that order) | | |

# List of References


Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2020). Rmarkdown: Dynamic documents for R. R package version 2.2.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and social psychology review*, *13*(3), 219–235.

Altman, D., & Royston, P. (2006). Statistics Notes: The cost of dichotomising continuous variables. *BMJ: British Medical Journal*, *332*(7549), 1080.

Altmann, G. (1999). Thematic role assignment in context. *Journal of Memory and Language*, *41*(1), 124–145.

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238. https://doi.org/10.1016/0010-0277(88)90020-0

Altmann, G., van Nice, K. Y., Garnham, A., & Henstra, J.-A. (1998). Late closure in context. *Journal of Memory and Language*, *38*(4), 459–484.

Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, *54*(3), 218–250. https://doi.org/10.1016/j.cogpsych.2006.07.001

Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations*. Manual.

Austin, P. C., & Brunner, L. J. (2004). Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses. *Statistics in medicine*, *23*(7), 1159–1178.

Bach, K. (2006). The top 10 misconceptions about implicature. *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn*, 21–30.

Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling Metaphor from Context: An ERP Study. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00559

Bambini, V., Canal, P., Resta, D., & Grimaldi, M. (2019). Time Course and Neurophysiological Underpinnings of Metaphor in Literary Context. *Discourse Processes*, *56*(1), 77–97. https://doi.org/10.1080/0163853X.2017.1401876

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211–227. https://doi.org/10.3758/BF03196968

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Blank, G. (1988). Metaphors in the lexicon. *Metaphor and Symbol*, *3*(3), 21–36.

Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 295–308. https://doi.org/10.1037/0278-7393.19.2.295

Boroditsky, L. (2000). Metaphoric Structuring: Understanding Time through Spatial Metaphors. *Cognition*, *75*, 1–28. https://doi.org/10.1016/S0010-0277(99)00073-6

Bowdle, B., & Gentner, D. (1997). Informativity and asymmetry in comparisons. *Cognitive Psychology*, *34*(3), 244–286.

Bowdle, B., & Gentner, D. (2005). The Career of Metaphor. *Psychological Review*, *112*(1), 193–216. https://doi.org/10.1037/0033-295X.112.1.193

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*(2), 211–252.

Brown, R. (1958). *Words and things.* The Free Press.

Cardillo, E. R., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior Research Methods*, *42*(3), 651–664.

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2012). From novel to familiar: Tuning the brain for metaphors. *Neuroimage*, *59*(4), 3212–3221. https://doi.org/10.1016/j.neuroimage.2011.11.079

Carston, R. (2010). XIII-Metaphor: Ad Hoc Concepts, Literal Meaning and Mental Images. *Proceedings of the Aristotelian Society (Hardback)*, *110*(3pt3), 295–321. https://doi.org/10.1111/j.1467-9264.2010.00288.x

Carston, R. (2012). Metaphor and the Literal–Nonliteral Distinction. In K. Allan & K. Jaszczolt (Eds.), *Cambridge Handbook of Pragmatics* (pp. 469–492). Cambridge University Press.

Chambers, C., Tanenhaus, M., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 687.

Chen, Z., & Cowan, N. (2005). Chunk limits and length limits in immediate recall: A reconciliation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1235.

Chiappe, D., & Kennedy, J. (1999). Aptness predicts preference for metaphors or similes, as well as recall bias. *Psychonomic Bulletin & Review*, *6*(4), 668–676. https://doi.org/10.3758/BF03212977

Chiappe, D., Kennedy, J. M., & Chiappe, P. (2003). Aptness is more important than comprehensibility in preference for metaphors and similes. *Poetics*, *31*(1), 51–68. https://doi.org/10.1016/S0304-422X(03)00003-2

Chiappe, D., Kennedy, J. M., & Smykowski, T. (2003). Reversibility, aptness, and the conventionality of metaphors and similes. *Metaphor and Symbol*, *18*(2), 85–105.

Chierchia, G., & McConnell-Ginet, S. (2000). *Meaning and Grammar: An Introduction to Semantics*. MIT Press.

Clark, H. H. (1996). *Using language*. New York, NY, US, Cambridge University Press. https://doi.org/10.2277/0521561582

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Clement, C. A., & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive science*, *15*(1), 89–132.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In *Eye movements: A window on mind and brain* (pp. 341–371). Amsterdam, Netherlands, Elsevier. https://doi.org/10.1016/B978-008044980-7/50017-3

Cohen, J. (1983). The cost of dichotomization. *Applied psychological measurement*, *7*(3), 249–253.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Cook, C., Heath, F., Thompson, R. L., & Thompson, B. (2001). Score reliability in webor internet-based surveys: Unnumbered graphic rating scales versus likert-type scales. *Educational and Psychological Measurement*, *61*(4), 697–706.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Coulson, S. (2012). Cognitive Neuroscience of Figurative Language. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge Handbook of Psycholinguistics* (First, pp. 523–538). Cambridge University Press. https://doi.org/10.1017/CBO9781139029377.027

Coulson, S., Davenport, T., Knoeferle, P., & Creel, S. C. (2015). Time Course of Metaphor Comprehension in the Visual World., In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*.

Coulson, S., & Lai, V. T. (Eds.). (2016). *The Metaphorical Brain*. Frontiers Media SA. https://doi.org/10.3389/978-2-88919-772-9

Coulson, S., & Oakley, T. (2005). Blending and coded meaning: Literal and figurative meaning in cognitive semantics. *Journal of Pragmatics*, *37*(10), 1510–1536. https://doi.org/10.1016/j.pragma.2004.09.010

Dahan, D., Tanenhaus, M., & Chambers, C. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2019). *Xtable: Export tables to LaTeX or HTML*. Manual.

Davidson, D. (2001). *Inquiries Into Truth and Interpretation: Philosophical Essays Volume 2*. Oxford University Press.

Dawson, N. V., & Weiss, R. (2012). Dichotomizing continuous variables in statistical analysis: A practice to avoid. *Medical Decision Making*, *32*(2), 225–226. https://doi.org/10.1177/0272989X12437605

Drummond, A. (2013). Ibex farm. *Online server: http://spellout. net/ibexfarm*.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye Movements as a Window into Real-Time Spoken Language Comprehension in

Natural Contexts. *Journal of Psycholinguistic Research*, *24*(6), 409–436. https://doi.org/10.1007/BF02143160

Elman, J. L. (2009). On the Meaning of Words and Dinosaur Bones: Lexical Knowledge Without a Lexicon. *Cognitive Science*, *33*(4), 547–582. https://doi.org/10.1111/j.1551-6709.2009.01023.x

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial intelligence*, *41*(1), 1–63.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science*, *22*(2), 133–187.

Fauconnier, G., & Turner, M. (2003). Conceptual blending, form and meaning. *Recherches en communication*, *19*(19), 57–86.

Fauconnier, G., & Turner, M. (2008). Rethinking Metaphor. In R. Gibbs (Ed.), *Cambridge Handbook of Metaphor and Thought*. Cambridge University Press, New York.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating Verbs, Situation Schemas, and Thematic Role Concepts. *Journal of Memory and Language*, *44*(4), 516–547. https://doi.org/10.1006/jmla.2000.2728

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press.

Fodor, J. A., & Lepore, E. (2002). *The compositionality papers*. Oxford University Press.

Forbus, K., Ferguson, R., & Gentner, D. (1994). Incremental structure-mapping, In *Proceedings of the Cognitive Science Society*.

Gagné, C. L. (2002). Metaphoric interpretations of comparison-based combinations. *Metaphor and Symbol*, *17*(3), 161–178.

Gambi, C., & Pickering, M. J. (2017). Models linking production and comprehension (E. M. Fernández & H. Smith Cairns, Eds.). *The Handbook of Psycholinguistics*, 157–181.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*(3), 181–215. https://doi.org/10.1016/0010-0277(94)90048-5

Gelman, A., & Hill, J. (2006). Data Analysis Using Regression And Multilevel/Hierarchical Models. In *Cambridge Universty Press*. https://doi.org/10.1017/CBO9780511790942

Gentner, D., & Boronat, C. (1992). Metaphor as mapping, In *Workshop on Metaphor, Tel Aviv, Israel*.

Gentner, D., & Bowdle, B. (2008). Metaphor as structure-mapping. In R. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 109–128). New York, NY, US, Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.008

Gentner, D., Bowdle, B. F., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In *The Analogical Mind: Perspectives from Cognitive Science* (pp. 199–253). Cambridge, MA, US, The MIT Press.

Gentner, D., & France, I. M. (1988). The verb mutability effect: Studies of the combinatorial semantics of nouns and verbs. In *Lexical Ambiguity Resolution* (pp. 343–382). Elsevier.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*(1), 45–56. https://doi.org/10.1037/0003-066X.52.1.45

Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language*, *37*(3), 331–355.

Gernsbacher, M. A., Keysar, B., Robertson, R. R., & Werner, N. K. (2001). The Role of Suppression and Enhancement in Understanding Metaphors. *Journal of Memory and Language*, *45*(3), 433–450. https://doi.org/10.1006/jmla.2000.2782

Gervais, W. M. (2020). Practical methodological reform needs good theory. *PsyArXiv*.

Gibbs, R. (1979). Contextual effects in understanding indirect requests. *Discourse Processes*, *2*(1), 1–10. https://doi.org/10.1080/01638537909544450

Gibbs, R. (1984). Literal Meaning and Psychological Theory. *Cognitive Science*, *8*(3), 275–304. https://doi.org/10.1016/S0364-0213(84)80004-X

Gibbs, R. (1986). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General*, *115*(1), 3–15. https://doi.org/10.1037/0096-3445.115.1.3

Gibbs, R. (1993). Process and products in making sense of tropes. In *Metaphor and Thought* (pp. 252–277). Cambridge University Press.

Gibbs, R. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, *34*(4), 457–486. https://doi.org/10.1016/S0378-2166(01)00046-7

Gibbs, R. (Ed.). (2008). *The Cambridge handbook of metaphor and thought*. New York, Cambridge University Press.

Gibbs, R., & Bryant, G. A. (2008). Striving for optimal relevance when answering questions. *Cognition*, *106*(1), 345–369.

Gildea, P., & Glucksberg, S. (1983). On understanding metaphor: The role of context. *Journal of Verbal Learning and Verbal Behavior*, *22*(5), 577–590.

Glucksberg, S. (1989). Metaphors in Conversation: How Are They Understood? Why Are They Used? *Metaphor and Symbolic Activity*, *4*(3), 125–143. https://doi.org/10.1207/s15327868ms0403_2

Glucksberg, S. (2001). *Understanding Figurative Language*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195111095.001.0001

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, *7*(2), 92–96. https://doi.org/10.1016/S1364-6613(02)00040-2

Glucksberg, S. (2008). How metaphors create categories–quickly. In R. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 67–83). New York, NY, US, Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.006

Glucksberg, S., & Estes, A. (2000). Feature accessibility in conceptual combination: Effects of context-induced relevance. *Psychonomic Bulletin & Review*, *7*(3), 510–515.

Glucksberg, S., Gildea, P., & Bookin, H. B. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning & Verbal Behavior*, *21*(1), 85–98. https://doi.org/10.1016/S0022-5371(82)90467-4

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, *97*(1), 3–18. https://doi.org/10.1037/0033-295X.97.1.3

Glucksberg, S., McGlone, M. S., & Manfredi, D. (1997). Property Attribution in Metaphor Comprehension. *Journal of Memory and Language*, *36*(1), 50–67. https://doi.org/10.1006/jmla.1996.2479

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge: Harvard University Press.

Guerra, E., & Knoeferle, P. (2014). Spatial distance effects on incremental semantic interpretation of abstract sentences: Evidence from eye tracking. *Cognition*, *133*(3), 535–552. https://doi.org/10.1016/j.cognition.2014.07.007

Guerra, E., & Knoeferle, P. (2017). Visually perceived spatial distance affects the interpretation of linguistically mediated social meaning during online language comprehension: An eye tracking reading study. *Journal of Memory and Language*, *92*, 43–56. https://doi.org/10.1016/j.jml.2016.05.004

Guerra, E., & Knoeferle, P. (2018). Semantic Interference and Facilitation: Understanding the Integration of Spatial Distance and Conceptual Similarity During Sentence Reading. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00718

Guest, O., & Martin, A. E. (2020). How computational modeling can force theory building in psychological science. *PsyArXiv*.

Hintz, F., Meyer, A. S., & Huettig, F. (2017). Predictors of verb-mediated anticipatory eye movements in the visual world. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(9), 1352–1374. https://doi.org/10.1037/xlm0000388

Hlavac, M. (2018). *Stargazer: Well-formatted regression and summary statistics tables.* Manual. Central European Labour Studies Institute (CELSI). Bratislava, Slovakia.

Holyoak, K. J., & Stamenković, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin*, *144*(6), 641–671. https://doi.org/10.1037/bul0000145

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415.

Huettig, F., & Altmann, G. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32. https://doi.org/10.1016/j.cognition.2004.10.003

Huettig, F., & Hartsuiker, R. J. (2008). When you name the pizza you look at the coin and the bread: Eye movements reveal semantic activation during word production. *Memory & Cognition*, *36*(2), 341–360. https://doi.org/10.3758/MC.36.2.341

Huettig, F., Mishra, R. K., & Olivers, C. N. L. (2012). Mechanisms and Representations of Language-Mediated Visual Attention. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00394

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Imbault, C., Shore, D., & Kuperman, V. (2018). Reliability of the sliding scale for collecting affective responses to words. *Behavior Research Methods*, *50*(6), 2399–2407.

Ivanko, S. L., & Pexman, P. M. (2003). Context Incongruity and Irony Processing. *Discourse Processes*, *35*(3), 241–279. https://doi.org/10.1207/S15326950DP3503_2

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford University Press.

Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 240.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, *118*(2), 126.

Johnson-Laird, P. N. (1987). The Mental Representation of the Meaning of Words. *Cognition*, *25*(1), 189–211. https://doi.org/10.1016/0010-0277(87)90009-6

Jones, L., & Estes, Z. (2005). Metaphor comprehension as attributive categorization. *Journal of Memory and Language*, *53*(1), 110–124. https://doi.org/10.1016/j.jml.2005.01.016

Jones, L., & Estes, Z. (2006). Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal of Memory and Language*, *55*(1), 18–32. https://doi.org/10.1016/j.jml.2006.02.004

Kamide, Y. (2008). Anticipatory Processes in Sentence Processing. *Language and Linguistics Compass*, *2*(4), 647–670. https://doi.org/10.1111/j.1749-818X.2008.00072.x

Kamide, Y., Altmann, G., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Keysar, B., Shen, Y., Glucksberg, S., & Horton, W. S. (2000). Conventional Language: How Metaphorical Is It? *Journal of Memory and Language*, *43*(4), 576–593. https://doi.org/10.1006/jmla.2000.2711

King, D., & Gentner, D. (2019). Polysemy and Verb Mutability: Differing Processes of Semantic Adjustment for Verbs and Nouns., In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, *7*(2), 257–266. https://doi.org/10.3758/BF03212981

Kirby, J. T. (1997). Aristotle on metaphor. *American Journal of Philology*, *118*(4), 517–554.

Knoeferle, P. (2015). Visually situated language comprehension in children and in adults. In R. K. Mishra, S. Narayanan, & F. Huettig (Eds.), *Attention and vision in language processing* (pp. 57–75). Springer.

Knoeferle, P. (2016). Characterising visual context effects. *Visually Situated Language Comprehension*, *93*, 227.

Knoeferle, P. (2019). Predicting (variability of) context effects in language comprehension. *Journal of Cultural Cognitive Science*, 1–18.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127. https://doi.org/10.1016/j.cognition.2004.03.002

Knoeferle, P., & Guerra, E. (2016). Visually Situated Language Comprehension. *Language and Linguistics Compass*, *10*(2), 66–82. https://doi.org/10.1111/lnc3.12177

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kusumi, T. (1987). Effects of categorical dissimilarity and affective similarity between constituent words on metaphor appreciation. *Journal of Psycholinguistic Research*, *16*(6), 577–595.

Lai, V. T., & Desai, R. H. (2016). The grounding of temporal metaphors. *Cortex*, *76*, 43–50. https://doi.org/10.1016/j.cortex.2015.12.007

Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research*, *1284*, 145–155. https://doi.org/10.1016/j.brainres.2009.05.088

Lakoff, G. (1990). The invariance hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics (includes Cognitive Linguistic Bibliography)*, *1*(1), 39–74.

Lakoff, G. (2008). The neural theory of metaphor. In R. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 17–38). New York, NY, US, Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.003

Lakoff, G., & Johnson, M. (2008). *Metaphors We Live By*. University of Chicago Press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*, 1–38.

Levin, S. R. (1982). Aristotle's theory of metaphor. *Philosophy & Rhetoric*, 24–46.

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, *79*, 328–348.

Lyngs, U. (2019). Oxforddown: An oxford university thesis template for r markdown. *GitHub repository*. https://doi.org/10.5281/zenodo.3484682

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19.

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843.

Mansfield, E. R., & Helms, B. P. (1982). Detecting Multicollinearity. *The American Statistician*, *36*(3a), 158–160. https://doi.org/10.1080/00031305.1982.10482818

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General*, *127*(4), 331–354. https://doi.org/10.1037/0096-3445.127.4.331

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.

McGlone, M. S., & Manfredi, D. A. (2001). Topic—vehicle interaction in metaphor comprehension. *Memory & Cognition*, *29*(8), 1209–1219. https://doi.org/10.3758/BF03206390

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, *66*(4), 545–567. https://doi.org/10.1016/j.jml.2012.01.001

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Müller, K. (2017). *Here: A simpler way to find your files*. Manual.

Murphy, G. L. (1996). On metaphoric representation. *Cognition*, *60*(2), 173–204. https://doi.org/10.1016/0010-0277(96)00711-1

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied Linear Statistical Models*. Irwin.

Newport, E., & Bellugi, U. (1978). Linguistic expression of category in a visual-gestural language: A flower is a flower is a flower. *Semantic factors in cognition*, 137–168.

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior*, *17*(4), 465–477. https://doi.org/10.1016/S0022-5371(78)90283-9

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056

Porretta, V., Kyröläinen, A., Van Rij, J., & Järvikivi, J. (2017). VWPre: Tools for preprocessing visual world data. *R package version*, *1*(0).

Pouscoulous, N., & Dulcinati, G. (2019). Metaphor. In C. Cummins & N. Katsos (Eds.), *Oxford Handbook of Experimental Semantics and Pragmatics*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198791768.013.19

Price, P. C., Jhangiani, R., & Chiang, I.-C. A. (2015). *Research methods in psychology*. BCCampus.

Pynte, J., Besson, M., Robichon, F.-H., & Poli, J. (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain and language*, *55*(3), 293–316.

R Core Team. (2020). *R: A language and environment for statistical computing*. Manual. R Foundation for Statistical Computing. Vienna, Austria.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372.

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Recanati, F. (2004). *Literal Meaning*. Cambridge University Press.

Rodriguez Ronderos, C., Guerra, E., & Knoeferle, P. (2020). The processing of German verb-object metaphors, In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*.

Rodriguez Ronderos, C., Noveck, I., & Tomlinson, J. (2020). Intentionality, speaker's attitude and the processing of verbal irony., In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*.

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in medicine*, *25*(1), 127–141.

RStudio Team. (2020). *RStudio: Integrated development environment for r*. Manual. RStudio, PBC. Boston, MA.

Rubio Fernandez, P. (2007). Suppression in Metaphor Interpretation: Differences between Meaning Selection and Meaning Construction. *Journal of Semantics*, *24*(4), 345–371. https://doi.org/10.1093/jos/ffm006

Rubio-Fernández, P., Cummins, C., & Tian, Y. (2016). Are single and extended metaphors processed differently? A test of two Relevance-Theoretic accounts. *Journal of Pragmatics*, *94*, 15–28. https://doi.org/10.1016/j.pragma.2016.01.005

Sato, M., Schafer, A. J., & Bergen, B. K. (2015). Metaphor priming in sentence production: Concrete pictures affect abstract language production. *Acta Psychologica*, *156*, 136–142. https://doi.org/10.1016/j.actpsy.2014.09.010

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 26.

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038.

Schmidt, G. L., Kranjec, A., Cardillo, E. R., & Chatterjee, A. (2010). Beyond Laterality: A Critical Assessment of Research on the Neural Basis of Metaphor. *Journal of the International Neuropsychological Society : JINS*, *16*(1), 1–5. https://doi.org/10.1017/S1355617709990543

Searle, J. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 76–116). New York, Cambridge University Press. https://doi.org/10.1017/CBO9780511609213.006

Sereno, S. C., Brewer, C. C., & O'Donnell, P. J. (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science*, *14*(4), 328–333.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition.* Oxford: Blackwell.

Sperber, D., & Wilson, D. (2008). A deflationary account of metaphors. In R. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 84–105). New York, NY, US, Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.007

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, *103*(2684), 677–680.

Swinney, D. A. (1979). Lexical Access during Sentence Comprehension: (Re)Consideration of Context Effects. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 645–659. https://doi.org/10.1016/S0022-5371(79)90355-4

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language*, *27*(3), 324.

Tabossi, P., & Zardon, F. (1993). Processing ambiguous words in context. *Journal of Memory and Language*, *32*(3), 359–372.

Tanenhaus, M., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of psycholinguistic research*, *29*(6), 557–580.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. C. (1996). Using eye movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation. In *Attention and performance 16: Information integration in perception and communication* (pp. 457–478). Cambridge, MA, US, The MIT Press.

Tanenhaus, M., & Trueswell, J. C. (2006). Eye movements and spoken language comprehension. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (pp. 863–900). Elsevier.

Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language*, *27*(6), 597–632.

Tendahl, M., & Gibbs, R. (2008). Complementary perspectives on metaphor: Cognitive linguistics and relevance theory. *Journal of Pragmatics*, *40*(11), 1823–1864. https://doi.org/10.1016/j.pragma.2008.02.001

Thibodeau, P., & Durgin, F. (2008). Productive figurative communication: Conventional metaphors facilitate the comprehension of related novel metaphors. *Journal of Memory and Language*, *58*(2), 521–540. https://doi.org/10.1016/j.jml.2007.05.001

Thibodeau, P., & Durgin, F. (2011). Metaphor Aptness and Conventionality: A Processing Fluency Account. *Metaphor and Symbol*, *26*(3), 206–226. https://doi.org/10.1080/10926488.2011.583196

Torreano, L. A., Cacciari, C., & Glucksberg, S. (2005). When Dogs Can Fly: Level of Abstraction as a Cue to Metaphorical Use of Verbs. *Metaphor and Symbol*, *20*(4), 259–274. https://doi.org/10.1207/s15327868ms2004_2

Trueswell, J., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic Influences On Parsing: Use of Thematic Role Information in Syntactic Ambiguity Resolution. *Journal of Memory and Language*, *33*(3), 285–318. https://doi.org/10.1006/jmla.1994.1014

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443.

van Rooij, I., & Baggio, G. (2020). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *PsyArXiv*.

Vasishth, S., Mertzen, D., Jäger, L. A., & Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language*, *103*, 151–175.

Vasishth, S., & Nicenboim, B. (2016). Statistical Methods for Linguistic Research: Foundational Ideas - Part I: Statistical Methods for Linguistics - Part I. *Language and Linguistics Compass*, *10*(8), 349–369. https://doi.org/10.1111/lnc3.12201

Vasishth, S., von der Malsburg, T., & Engelmann, F. (2013). What eye movements can tell us about sentence comprehension: Eye movements and sentence comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 125–134. https://doi.org/10.1002/wcs.1209

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth). New York, Springer.

von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, *94*, 119–133. https://doi.org/10.1016/j.jml.2016.10.003

Warriner, A. B., Shore, D. I., Schmidt, L. A., Imbault, C. L., & Kuperman, V. (2017). Sliding into happiness: A new tool for measuring affective responses to words. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *71*(1), 71.

Weiland, H., Bambini, V., & Schumacher, P. B. (2014). The role of literal meaning in figurative language comprehension: Evidence from masked priming ERP. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00583

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Wickham, H., François, R., Henry, L., & Müller, K. (2020). *Dplyr: A grammar of data manipulation*. Manual.

Wickham, H., & Henry, L. (2020). *Tidyr: Tidy messy data.* Manual.

Wilson, D. (2003). Relevance and lexical pragmatics. *Italian Journal of Linguistics*, *15*, 273–292.

Wilson, D. (2011). Parallels and Differences in the Treatment of Metaphor in Relevance Theory and Cognitive Linguistics. *Intercultural Pragmatics*, *8*(2), 177–196. https://doi.org/10.1515/iprg.2011.009

Wilson, D., & Carston, R. (2006). Metaphor, Relevance and the ’Emergent Property’ Issue. *Mind $< $html_ent glyph="@amp;" ascii="&"/$> $ Language*, *21*(3), 404–433. https://doi.org/10.1111/j.1468-0017.2006.00284.x

Wilson, D., & Carston, R. (2007). A unitary approach to Lexical Pragmatics: Relevance, Inference and Ad hoc concepts. In N. Burton-Roberts (Ed.), *Pragmatics* (pp. 230–259). London, Palgrave Macmillan UK. https://doi.org/10.1057/978-1-349-73908-0_12

Wilson, D., & Sperber, D. (1999). Relevance and relevance theory, In *MIT Encyclopedia of the Cognitive Sciences*, Citeseer.

Wilson, D., & Sperber, D. (2004). Relevance theory. *Handbook of Pragmatics. Oxford: Blackwell*, 607–632.

Wilson, D., & Sperber, D. (2012). *Meaning and Relevance.* Cambridge, Cambridge University Press. https://doi.org/10.1017/CBO9781139028370

Wolff, P., & Gentner, D. (2000). Evidence for role-neutral initial processing of metaphors. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*(2), 529–541. https://doi.org/10.1037//0278-7393.26.2.529

Wolff, P., & Gentner, D. (2011). Structure-Mapping in Metaphor Comprehension. *Cognitive Science*, *35*(8), 1456–1488. https://doi.org/10.1111/j.1551-6709.2011.01194.x

Xie, Y. (2015). *Dynamic Documents with R and knitr* (Vol. 29). CRC Press.

York, R. (2012). Residualization is not the answer: Rethinking how to address multicollinearity. *Social science research*, *41*(6), 1379–1386.

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832